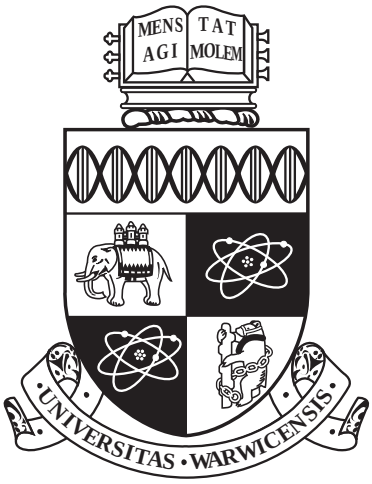


Point Process Survival Models for Epilepsy Data

by

Boryana Cristina Lopez Kolkovska

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Statistics

June 2016

THE UNIVERSITY OF
WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my supervisor, Professor Jane L. Hutton, for her valuable guidance, support and encouragement throughout the entire duration of this thesis. Her energy and hard work have always been a source of inspiration for me and the people around her.

Special thanks go to Professor Mark F. J. Steel, Doctor Simon Spencer and Doctor Julia Brettschneider for their careful consideration of the thesis and their thoughtful suggestions to improve some areas of research in this work.

I thank CONACyT (The National Council for Science and Technology, Mexico, CVU No. 269461) for the financial support granted through my research studies.

My deep gratitude to my parents, who have always been my pillars and live examples for me to follow in every aspect. To my brothers: Alejandro, for his firm points of view and guidance, Alfredo for his love and character, and Rodrigo, who knows how to bring out a smile in the toughest of times.

To my love, Eduardo Martínez Sosa, who has always walked by my side regardless of distance, and through whose love and strength I have become a better person.

To Suzanne Higham for her constant friendship and welcoming heart, and to Raina Wilson, who inspires kindness in those she meets. And to Tug, for ever our beloved Tug.

Declarations

I hereby declare that this is my own work, except where explicitly stated, and that it has not been submitted for a degree at any other university.

All the chapters in this thesis are joint work with my supervisor Prof. Jane L. Hutton.

Abstract

The work carried out in this thesis is focused on the proposal, comparison and assessment of survival analysis models for epilepsy data. Although the Cox proportional hazards model provides a popular approach to medical recurrent events modelling, other accelerated life alternatives seem more appropriate when compared under goodness of fit tests. In our research we apply the Cox proportional hazards model and two models consisting of a Poisson-Gamma mixture model that could assume the existence of a cure fraction , and which have been developed and proposed by B. Cowling[11] and J. Rogers[39] respectively. We applied these methods to the Multicentre study of early Epilepsy and Single Seizures (MESS) data set. In this epilepsy study, patients with different types of seizures were randomized to either immediate or delayed treatment, which consisted in being administered one of seven types of drugs. The aim of the study consisted in producing a prognosis with which the clinicians and patients could take an informed decision on whether or not it was preferable to take an anti-epileptic drug.

We investigated the behaviour of the survival function for the Cox proportional hazards model, the joint model and the joint with cure fraction model under the epilepsy data set, under the framework of residual analysis studies, as well as empirical vs theoretical survival functions.

As a final contribution of our work, we proposed modification of the accelerated life models. Since a patient cannot be diagnosed with epilepsy unless he or she presents at least two un-provoked seizures, we proposed a zero-truncated joint model, which considers the pre-randomization counts to be strictly positive. This model has been extended to consider a cure fraction of the population, but is still under development, since the corresponding parameter estimations become considerably more complicated.

Chapter 1

Introduction

The main motivation of this thesis lies in the study of individual patient epilepsy data under the framework of point process survival analysis, which includes the seizure count pre-randomization, seizure recurrence times post-randomization, the treatment allocation at randomization and individual baseline data such as age, sex and electroencephalogram outcomes. The patients belonging to this study and their clinicians are interested in obtaining a scheme which allows them to take an informed decision on whether or not it is most convenient for the patient to take a certain type of drug, depending on the characteristics of his or her condition. The common approach to this kind of medical study in survival analysis considers the occurrence count as part of the covariates for the model, and provides a population prognostic where individual seizure recurrence is not considered. For researchers who are working with epilepsy for the first time, the book by Marson [35] introduces the causes and characteristics of epilepsy with accuracy and simplicity, based in some measure on the proper terminology by the International League Against Epilepsy (ILAE), as described in the paper by Engel[15].

We are interested in taking a statistical approach where the information provided from the seizure count and the time to event can be used to jointly model the seizure recurrence probability for individual patients, and in this thesis a variety

of existing and newly proposed point process survival models are presented, analyzed and contrasted between one another in an attempt to provide a more informative prognostic for this condition.

In the first section of this chapter we provide a brief literature review about survival and residual analysis, followed by a concise overview of the chapters that comprise this thesis.

1.1 Analysis of Survival Analysis

When considering the modeling of epilepsy data as a count data problem, a logical approach is to use the Poisson process generalized linear model(McCullagh and Nelder [36]). However, due to the overdispersion usually observed in lifetime data such as epilepsy data, the negative binomial distribution for the seizure counts is a more appropriate model, as discussed in Greenwood & Yule[19]. Hougaard [23] presents alternative count process distributions for data with overdispersion, such as the inverse Gaussian distribution.

The literature about the theory and applications for modeling recurrent events is very extensive. As we will introduce later on more extensively, the models that are dedicated to study the time from a specific start at which a random unit or individual from a population will experience a recurrent event are known as "Live distributions", "Reliability Models", or as they are more commonly known in life sciences, "Survival Models". An extense discussion and introduction to these kinds of models can be found in the books by Meeker and Escobar [37], Lawless [31], Leemis [32], Tobias and Trindade [44] and Collett [7].

The first epilepsy model we study for the epilepsy MESS data was proposed by Kim et al. [27], which consists of a Cox proportional hazards model under a set of specific covariates. Traditionally lifetime data such as epilepsy recurrence are studied under proportional hazards models. The propriety of such usage of the

model has been questioned before, for example in the case of cerebral palsy studies (Kwong & Hutton[30]). This conclusion first arose from the work by Hutton & Solomon ([25]), in which they use the mixture model proposed by Cox & Reid ([12]) to study mixture models of proportional hazards and accelerated life models.

Cowling, Hutton & Shaw ([11]) have proposed a Poisson-gamma mixture model, which is discussed in more detail in the thesis by Cowling [10]. The model in question proposes an individual seizure rate, where pre-randomization counts and post-randomization seizure times are considered as variables, unlike previous works in which the pre-randomization counts tend to be used only as covariates. Rogers([39],[41]) has generalized the model by Cowling by assuming that there is a proportion of the population who attains remission before the study finishes, or in other words, patients who do not experience a seizure recurrence after randomization. For both models we modify the pre-randomization assumptions, since we know that in practice all patients must have at least one seizure pre-randomization. This assumption leads to a zero-truncated model, where the likelihood and score functions present significant changes. For the theory behind truncated models, the literature offers comprehensive discussions (Klein & Moeschberger[28], Cohen & Clifford[6]).

1.2 Goodness of Fit

As part of this work, we consider the Cox proportional hazards model proposed by Kim et al., the joint model proposed by Cowling and the joint model with cure fraction by Rogers. We have compared the model predictions on the epilepsy data, after which we carried on to study under a residual analysis for goodness of fit.

Residual analysis is a goodness-of-fit tool frequently used for the generalized linear model regressions (McCullagh & Nelder[36]), but there is a limited knowledge of the proper interpretation and behaviour of residual analysis for survival models. This aspect is extensively discussed by Collett[7], who argues that the residual

shapes for bad fits are commonly known, but there is no clear expectation of how residuals should appear under a good model fit.

1.3 Overview of thesis

We formally present the fundamental definitions characterizing epilepsy in Chapter 2, where we endeavour to provide an intuitive understanding of the types of epilepsy seizures depending on their onset location and the spreading behavior of the abnormal electrical impulse in the brain. The second section of this chapter consists of the introduction to the main object of study in the thesis, the MESS epilepsy data which consists of an individual patient baseline and progression data throughout a treatment scheme which allocates the individuals randomly either to immediate or to deferred treatment. The second epilepsy data is based on the SANAD study, which is devoted to compare drug efficacy between the typical drug (Carbamazepine for partial seizures, and Valproate for general seizures), and a group of alternative drugs with the aim of finding the best drug for each type of treatment. The final part of this chapter is dedicated to the introduction to the fundamental theory of survival analysis, likelihood theory and the most common parametric survival models.

In Chapter 3 we present a chronological review of the relevant theory behind the modeling of medical survival data. The first section includes a discussion about the robustness of regression estimates under survival model misspecification, and in particular, the contrast between proportional hazards models and accelerated life models for epilepsy, under orthogonal parametrization. The second section presents an evolution of the survival models that have been proposed for seizure recurrence in epilepsy, where we reproduce the results under the survival model proposed by Kim et al[27], the first paper where the MESS data set has been modeled. We aim with this studies to lead the reader to the final works in survival analysis under consideration, where the seizure counts and the times to recurrence are modeled

jointly under a point process framework. These models, proposed by Cowling[10] and Rogers[39] have been the focus of our attention for the majority of the thesis research time, and were the object of study of the fifth chapter, which we will describe below.

From the fitted values analysis under the point process survival models, we propose in Chapter 4 a zero-truncation of the seizure count density due to the inherent nature of epilepsy. Given that a patient cannot be diagnosed with epilepsy unless he has experienced at least two occurrences, the assumption of at least one pre-randomization seizure should provide a better approximation to the phenomenon of interest. The seizure count and seizure recurrence time joint model is presented, along with the corresponding marginal distributions, the log-likelihood and its derivatives with respect to the parameters of interest. We fit and compare this model with its non-truncated analogous joint model. In the second section of this chapter, we introduce a generalization of the joint model by considering a dependent frailty term, which is characterized by a set of covariates. This generalization is still under development, but the main interest lies in that the inclusion of covariates can produce a more specific measure of the level of heterogeneity between sub-populations of patients with a particular set of symptoms. In the last section of the chapter, we develop a zero-truncated version of the model proposed by Rogers, where the joint density, its likelihood and derivatives with respect to the parameters of interest are shown. Further work is needed to obtain the proper numerical approximation.

In Chapter 5 of the thesis is dedicated to the goodness of fit study of, mainly, the Cox proportional hazards model, the joint model proposed by Cowling and the joint model with a cure fraction proposed by Rogers. The large majority of this chapter presents a contrast between these models under a residual analysis, where we aim at finding an improvement on the seizure recurrence prediction when considering the seizure count as a random variable rather than a covariate. We

begin by introducing the residual score definitions, and proceed to compare the corresponding values of the epilepsy residuals under the models under consideration. Given the variety of results and interpretations of the results under the data, we carried out a simulation study under the point process models, and discuss the parameter optimization sensitivity for each case. The second section of this chapter displays a series of plots contrasting the Kaplan-Meier survival curves against the theoretical marginal time to seizure survival functions, where we aim to compare the similarities and differences between the model predictions, as well as observe possible reasons for the difference in the model fits. The majority of the model diagnostics carried out in this chapter are focused on the joint models proposed by Cowling and Rogers due to the chronological order in which the research was carried, but we are interested on reproducing the residual analysis and the survival function curves in a further work.

Chapter 2

Introduction to Epilepsy Data and Survival Analysis

In the following chapter we aim to introduce the reader to relevant statistical model concepts and the classification and causes for epilepsy. Later on we present a description of the MESS epilepsy data set, which will be our object of study, followed by a quick introduction to the SANAD epilepsy data base. Finally we present the relevant survival analysis preliminaries from which the models under study have been developed.

2.1 Introduction to Epilepsy Disease

The presence of epilepsy is a clinical diagnosis which should always involve a series of tests for the presence of epileptic seizures. The classification of types of epilepsy has proved to be such a colossal task, that there exist research and medical groups dedicated to this purpose. Since 1997 the International League against Epilepsy (ILAE) Task Force on Classification and Terminology has been dedicated to the evaluation of the pertinent terms and aspects of the classification of epilepsies. We present here a brief introduction to the terminologies and characteristics of epilepsy,

based on the works by the ILAE ([15]) and the book by Appleton and Marson ([3]).

An epileptic seizure (also called and ictal event) is an abnormal electrical discharge in the brain cortex, which consists of the layers of neuron cells laying on the surface of the brain. These abnormal electrical discharges are caused when a number of damaged cells fail to communicate properly to neighboring neurons. Because of the lack of communication, the neurons situated next to the damaged cells are affected by becoming over-excitable, and produce discharges in a chain reaction to other neurons, thus causing the seizure. The types of seizures vary depending on the starting location of the discharge in the cortex.

The large variety of types of seizures makes it difficult to define epilepsy in a such a broad meaning that it includes all known types of recurrent seizures. Observe however that the existence of a seizure is not in itself diagnosed as epilepsy. A person is said to have epilepsy when they experience two or more unprovoked epileptic seizures, generally during a reasonable span of time. A practitioner might doubt to diagnose a patient as epileptic if he is known to have experience one epileptic seizure at five years old, and a second one at seventy-five years old.

Depending on the onset location of the electric discharge in the brain, and how such discharge travels through the cortex, three types of epilepsy branches are defined. If a discharge commences in one of the two cerebral lobes and during expansion remains in the same lobe, then this is said to be a partial or focal seizure. If the electrical discharge originates on one lobe, spreads towards the center link of the brain and spreads to the two lobes, it is said to be a partial or focal seizure with secondary generalization. When the seizure begins at the center of the brain cortex, it expands to the two lobes and then it is called a generalized or primary seizure. These three possible types of epilepsy branches stem several epilepsy types:

- *Tonic-Clonic seizures*: These seizures appear as a result of a generalized epilepsy or partial with second generalized epilepsy, and consist of two phases. The first phase is the tonic (contraction) phase, which typically lasts one to

two minutes. During this phase the person experiences contraction of his muscles over all the body, and because of the loss of control over the muscles, the patient loses his equilibrium and fall to the ground in a rigid way. Because the muscles in the trunk are also contracting, this forces the air out of the lungs and breathing becomes uncoordinated. The sustained contraction of the muscles causes the oxygen in the body to be quickly used up. Clenched jaws may cause the patient to bite his tongue or cheeks, and the contractions may cause incontinence of urine or faeces. The second phase of the seizures is the clonic (convulsive) phase, which tends to last from two to three minutes. During this period the individual experiences rhythmic bodily jerks, increasing during the first one or two minutes before they diminish and cease, leaving the person lying unconscious. When the person regains consciousness, he has to be helped to sit down and often feels confused, tired and experiences head aches.

- *Myoclonic seizures*: these are bodily jerks, often shock-like contractions of the limbs.

- *Typical Absence seizures*: These types of seizures last usually from five to ten seconds, and characterized by a specific type of EEG (electroencephalogram) reading. This disease is almost always found in children and the symptoms may not be noticed by the parents or the child himself. It occurs suddenly, making the person stop talking or doing something in a paused manner, until he resumes his activities, often not realizing what has happened. With some frequency these absence seizures are followed by short-lived myoclonic seizures.

- *Focal or Partial seizures*: these types of seizures are distinguished to have two main manifestations, depending on the site of the cortex where they affect the neurons. When the seizure occurs in the motor area of the cortex, there is an external manifestation of the brain insult. Since the most excitable neu-

rons in the motor cortex correspond to the thumb, index finger, mouth corner and the toe, the patient experiences bodily jerks usually in such areas, but not restricted to them, and in opposite direction of the affected lobe in the brain. Sometimes the patient starts to turn his head and eyes in the direction of the damaged lobe (versive seizures) or in the opposite direction (adversive seizures). In these cases, the body reacts to the seizure but the consciousness of the person remains unchanged, he is aware of what is happening and his perception remains intact. In these cases, when the consciousness is not affected, the seizures are called *"simple partial seizures"*. The second type of focal or partial seizures occur when the parietal, temporal and occipital lobes are affected, just behind the motor cortex area. These types of seizures cause a disturbance of internal rather than external perception, although the effects depend once more on the area of the brain that is being affected. When the seizure occurs in the parietal lobe, the person can experience the feeling known as "pins and needles", numbness or heaviness of the limbs. When the occipital lobe is being affected, simple visual hallucinations are experienced such as seeing colored dots or geometric shapes. Finally, when the temporal lobe experiences a seizure, smell and taste hallucinations, often unpleasant, and inexplicable fears are experienced when the anterior part of the lobe is affected. When the posterior lobe is affected the person experiences complex visual hallucinations, such as staring at himself from afar. Other possible reactions can include emotional distress (such as sudden fear or panic attacks), vertigo, chest and abdominal sensations, the feeling of something being previously experienced (known as "dej vu"), loss of memory (amnesia) or automatic behaviours (such as putting clothes on and off repetitively). When the dominant lobe of the brain in a person is disturbed by a seizure, speech, comprehension and communication abilities are affected. Partial seizures in which the consciousness is affected or disturbed are usually called *"complex*

partial seizures". Some of these partial seizures sometimes can expand to the center of the brain and then to the other lobe, hence becoming from a second generalized to a tonic-clonic seizure.

An epilepsy syndrome is not defined solely by the type of epilepsy that a person experiences. An epilepsy syndrome is defined by a cluster of characteristics ranging from EEG readings to clinical features such as the age when the seizures started, the type or types of seizures, family history when possible and developmental and neurological findings. The identification of an epilepsy syndrome helps to identify which antiepileptic drugs are most convenient, and to elucidate the possible prognosis in a patient. The prognosis means that the clinician will be able to give information about the possible cause of the seizures, if the person may develop difficulties such as learning impairments and if the epilepsy is likely to be controlled by drugs or to disappear spontaneously.

2.2 Epilepsy Data

The main interest of this work is to study and model the process that leads to seizure recurrence for patients presenting epileptic seizures, and to provide a prognostic analysis in order to assist the patients and their clinicians on deciding which epilepsy treatment to follow, if need be. The object of study focuses on a multicentre trial that follows-up 1443 patients with different types of Epilepsy. The data collected before, at and after randomization of the patients to two treatment schemes is presented in a summarized fashion.

During the study of relevant covariates (such as age and type of epilepsy), we were able to fit the models discussed in later chapters to another important epilepsy data set named SANAD. Such study aimed to observe the patient's response to two main types of drugs (Carbamazepine and Valproate), and their respective alternative drugs. For this reason SANAD has two separate sub-studies or Arms, called Arm

A for studies for Carbamazepine and Arm B for studies with Valproate.

Mess Study introduction

The Multicentre study of early Epilepsy and Single Seizures (MESS) consists of a study and its resulting data set. Patients in the study had a history of epileptic seizures and their clinicians were unsure of the need for an anti-epileptic drug (AED). Patients were recruited for over five years, and randomized to one of two policies: deferred or immediate treatment. The aims of this study are:

- to measure the differences in policies
- to define prognostic factors for seizure recurrence
- to define psychosocial outcomes of the policies, and
- to make results available in a form which allows patients to make informed decisions.

When treating with AEDs, we expect certain advantages such as the reduction of the number of seizures in the short term, and the remission from epilepsy in the future. However, from the use of such AEDs the patients risk physical and psychological damage, sometimes serious. This study developed as an attempt to assess which policies are optimal for the diverse groups of epilepsy patients. It is of particular interest to understand what the risk of recurrence is, once a first seizure has happened, and how the treatment alters that risk.

The patients were eligible for the trial if they presented with a history of one or more spontaneous, unprovoked epilepsy seizures, were at least one month old, had not taken AEDs before, and they were, along with their clinicians, uncertain of needing treatment. Patients with a progressive disease were not considered for the study.

MESS randomized patients with single seizures, subjects with infrequent seizures and patients with more frequent seizures with less severe symptomatology. Thus, the MESS study randomized patients presenting tonic-clonic seizures pre-randomization (including primary and secondary generalized seizures), and with seizures that were simple partial, complex partial, or generalized absence and myoclonic seizures.

Between 1993 and 2000, 1847 patients were invited to join the trial, from UK and non-UK medical centres. Of these 1847 patients, 404 did not consent to randomization, while subsequently 722 were randomly assigned to immediate treatment, and 721 to deferred treatment. Over five years a total of 1443 patients were recruited. The randomization process was undertaken by an independent randomization centre, that balanced across two factors: centre or region, and number of seizures at randomization. The recruitement resulted in 56% of the patients having a single seizure in their history, and the remaining 44% had a history of at least two epileptic seizures. There were 717 (49.6%) subjects from the UK, and 726 (50.3%) from other countries, as shown in Table 2.1.

For patients allocated to the immediate policy, their clinicians determined the optimal AED for the patient and prescribed it as soon as possible. Most of the patients in this group were treated with Carbamazepine (CBZ, 45%), or with Valproate (VPS, 45%) and 10% with other AEDs. The remaining patients in the deferred treatment policy received no drugs until the clinician and patient determined that treatment was necessary. Of such deferred-located patients, 332 started treatment during the course of the trial: 134 (40%) with carbamazepine, 142 (43%) with valproate, and the remaining 37 with another AED.

Of the patient's history pre-randomization, data such as date of birth, the times to seizures and total number of seizures prior to randomization, and the type of epilepsy for each individual was collected. The information gathered at randomization included imaging, Electro-encephalogram (EEG) outcome if it was available,

	Immediate trt. ($n = 722$)		Deferred trt. ($n = 721$)	
	Number	%	Number	%
Sex				
Male	403	56%	423	58%
Female	319	44%	298	42%
Centre				
UK	363	50%	354	49%
Non-UK	359	50%	367	51%
EEG abnormalities				
Non-specific abnormality only	83	11%	88	12%
Generalized	131	18%	105	15%
Focal	184	25%	200	28%
Imaging abnormal	71	10%	69	10%
Seizure types pre-randomization				
Simple partial	15	2%	20	3%
Complex partial	36	5%	32	4%
Secondary generalized Tonic-Clonic	239	33%	215	30%
Myoclonus only	6	$< 1\%$	5	$< 1\%$
Absence only	3	$< 1\%$	3	$< 1\%$
Tonic-Clonic seizures	375	52%	406	56%
Combinations of generalized seizures	21	3%	19	3%
Other seizures	17	2%	13	2%
No. of seizures pre-randomization				
1	404	56%	408	57%
2	183	25%	165	23%
3	50	7%	58	8%
4	28	4%	18	2%
$5 - 9$	30	4%	36	5%
≥ 10	17	2%	28	4%
Clinical and family history				
Developmental delay/learning disability	34	5%	23	3%
Neurological deficit	52	7%	40	6%
Previous neurological insult	99	14%	90	12%
Previous febrile seizures	53	7%	52	7%
Previous acute symptomatic seizures	14	2%	19	3%
First-degree family history of seizures	76	11%	86	12%

Table 2.1: MESS study data outcome, tabulated by patients allocated to immediate and deferred treatment respectively.

age at randomization, date of randomization. The types of EEG abnormalities, the allocation to treatment policy, type and dose of drug were also recorded. The patients' follow-up lasted up to 8 years. The outcomes were the times to first, second and fifth seizures post-randomization and the types of seizures. This information is summarized in Table 2.1, which extracted from Marson et al [35].

Since with the passing of time, patients in the trial died or left the study, it can be observed that the proportion of patients receiving an AED in the two groups becomes gradually smaller. At 5 years from randomization, 60% of the patients in the immediate policy group are still receiving treatment, contrasted with 41% in the deferred policy group.

Due to the small number of patients presenting with myoclonic, absence, T-C with generalized or other seizures, these categories were not considered for the statistical analysis. Additionally, five individuals were taken from the sample, for whom there was no information of their pre-randomization history. These reductions resulted in a sample size of 1334 patients. The following tables and plots are based on these 1334 patients.

In the histogram presented in Figure 2.1, the proportion of patients that received a type of drug at a certain age is shown. Observe that the majority of patients are children and young adults from 8 to 31 years old, and that the main AEDs administered were carbamazepine and valproate, shown in red and yellow respectively. Indeed, in table 2.2 it can be seen that, from the 659 patients receiving an AED, 316 (47.9%) are treated with carbamazepine and 290 (44.0%) with valproate. Only 1 person has been treated with oxcarbazepine and 3 more with vigabatrin. The difference in drug selection is to be expected as valproate is typically a first line AED for patients presenting with generalized seizures, whilst carbamazepine is usually the corresponding first line AED for individuals presenting with partial seizures.

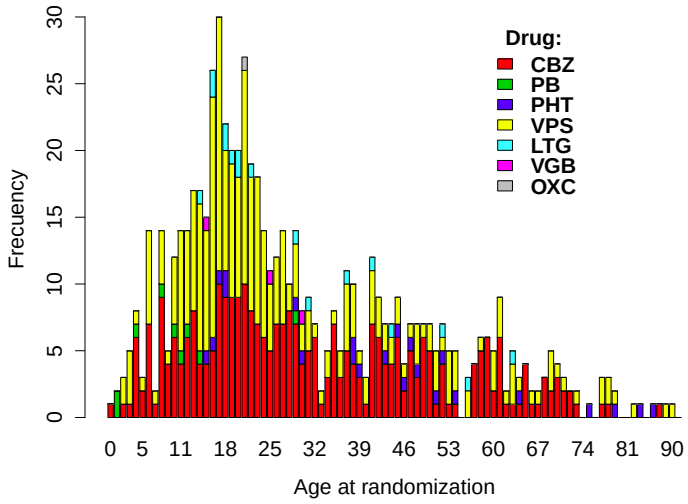


Figure 2.1: Age of patients at randomization

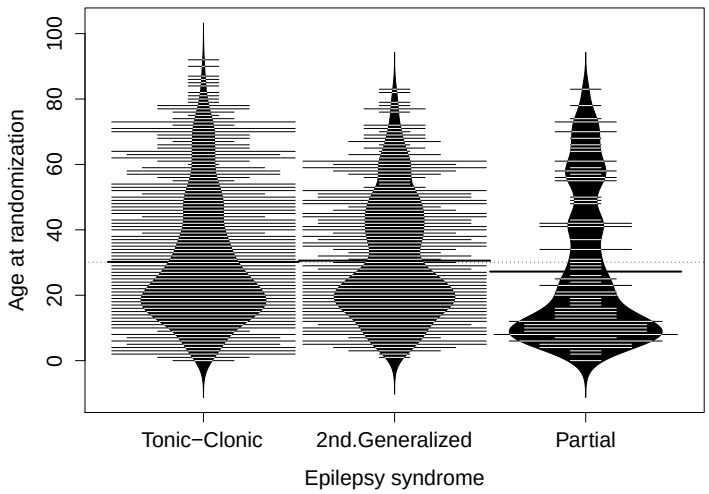


Figure 2.2: Age of patients at randomization by epilepsy syndrome.

Drug	n	Mean age at randomization		% Female	% Partial	% 2nd.General	% T-C
CBZ	316	31.79	(18.7)	44.6	39.5	50.3	10.1
PB	9	10	(8.5)	77.7	0	77.7	22.2
PHT	23	44.7	(21.5)	34.7	4.3	21.7	73.9
VPS	290	27.35	(18.2)	42.0	4.4	22.0	73.4
LTG	17	30.35	(15.6)	70.5	17.6	29.4	52.9
VGB	3	23.33	(7.6)	33.3	33.3	66.6	0
OXC	1	21	–	100	0	100	0
Total	659	29.9	(18.8)	44.3	7.5	36.1	56.3

Table 2.2: MESS study data outcome, tabulating the patients receiving immediate treatment by the drug type received.

The average age at randomization of the patients in the subgroup of individuals who were treated is consistent to the mean age at randomization of the whole population, as can be observed from the bean plots in Figure 2.2. Bean plots are produced from the R package *'beanplot'* [26], and present an alternative to boxplots by presenting lines as scatter plots and thus allowing the viewer to identify overlapping values. The second useful feature of such plots is that the empirical density is presented and the average is shown both for the whole population and for the subgroups. In our case, the plot shows that the density of patients with Tonic-Clonic and second generalized seizures are concentrated around 20 years old although both share a mean of 30 years old according to the mean of the entire population. For patients with partial seizures only, the majority of individuals are children and the number of observations is visibly lower than for the other two types of epilepsy.

Partial seizures tend to be more frequent in children and to manifest more often than tonic-clonic or second generalized seizures. In Table 2.3 it is shown that before randomization, of the three types of epilepsies considered, half of the patients with partial seizures had two seizures, which for the other two types of seizures half the population only had one. In contrast, there are patients with second generalized seizures who experience up to 130 seizures prior to being randomized in the study. Patients with partial seizures had in average almost twice the number

Epilepsy type	n	Pre-randomization count				
		mean	s.d.	median	min.	max.
Partial	103	6.049	13.61	2	1	99
2nd.General	453	3.366	10.76	1	1	130
Tonic-Clonic	778	1.515	0.99	1	1	10
Total	1334	2.494	7.47	1	1	130

Table 2.3: In this table we show the number of patients corresponding to each epilepsy syndrome, and for each group, the general distribution of the number of seizures pre-randomization is shown.

of occurrences than patients with second generalization, and four times more, in average, than individuals with tonic-clonic seizures only. However, the number of recruited patients with tonic-clonic and second generalized seizures are significally greater than those with partial seizures.

In terms of the times of first seizures post-randomization, shown in Table 2.4, the seizure recurrence seems consistent in the sense that patients with partial seizures who were treated with CBZ have shown a noticeable improvement over patients treated with VPS. No patient with partial epilepsy under CBZ treatment has had a seizure before 13 days from the randomization, as opposed to any other group under treatment, and by the time half the population had had a seizure, it was 2.7 times longer than the same time observed for the population with partial epilepsy under VPS. For the population of patients with Tonic-Clonic seizures only, the effect of the two drugs at a simple glance is not apparent. For the patients with second generalized epilepsy, however, the mean and median indicate an improvement of those patients treated with VPS over those under CBZ.

SANAD study

The SANAD study is an unblinded randomized controlled trial in hospital-based clinics in the UK. It is divided into two arms, which compared the effects of new antiepileptic drugs to the standard drugs generally used. Arm A consisted of the

Epilepsy type	Drug	% obs.	Days to first post-randomization seizure mean	s.d.	median	min	max
Partial	CBZ	43.75	856.10	885.92	503	13	2825
Partial	VPS	38.47	674.7	977.76	185	1	2922
2nd.General	CBZ	57.86	929.6	757.23	932	1	2884
2nd.General	VPS	56.25	1231.0	983.04	1230	1	2816
Tonic-Clonic	CBZ	64.80	1184.0	928.74	1183	1	3127
Tonic-Clonic	VPS	55.86	1090.0	904.47	1072	1	3161
	Total	54.81	1036.0	876.47	975.5	1	3161

% obs. denotes the proportion of observed seizure times in the group.

Table 2.4: In this table we show the number of patients corresponding to each epilepsy syndrome, and their first seizure post-randomization distribution for those patients in each group receiving Carbamazepine or Valproate as the treatment drug.

study comparing gabapentin (GBP), lamotrigine (LTG), oxcarbazepine (OXC) and topiramate (TPM) to carbamazepine (CBZ), in which 1721 patients were recruited. Arm B compared the drugs lamotrigine and topiramate to the typical drug valproate (VPS), and recruited 716 patients to be studied under this scheme. The study outcomes are the time to treatment failure and the time to 12 months remission from seizures. In the analysis presented below, we use the times to treatment failure overall as the time to event. The data set we are working with consists of the observations of 959 patients with epilepsy, 737 of which are in Arm A, and the remaining 213 are in Arm B. This subset of the entire SANAD data base consists of the patients in the study whose DNA sequence is currently being obtained and added to the data, with the purpose of studying in the future the correlation between the epilepsy types and drug responses, with their genetic possible predisposition. As can be observed in Table 2.5, the proportions of epilepsy seizure types (Generalized vs Partial) differ across the two arms studied. While Arm A has a great majority of patients with partial seizures, Arm B directs more attention to patients with generalized seizures, having about the same number of unclassified cases.

The difference in epilepsy type proportions can be observed clearly in Figure 2.3, where the Kaplan-Meier curves for the estimated survival curves are shown.

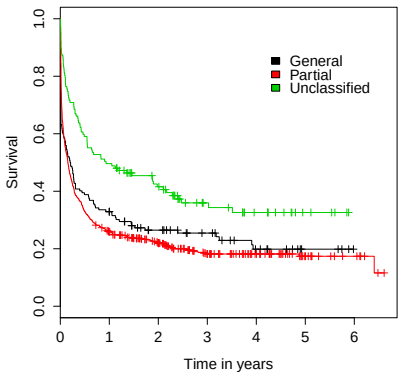
	General	Partial	Unclassified
Full data set	152	677	127
Arm A	8	662	64
Arm B	144	15	63

Table 2.5: Number of patients by syndrome and by Arm.

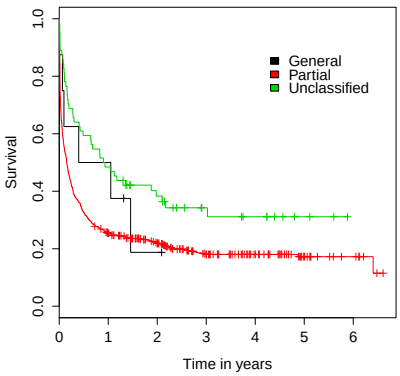
In the first plot we observe the survival curves corresponding to the epilepsy type, without specifying the Arm factor. The unclassified-epilepsy patients seem to have a higher probability of surviving than the patients presenting generalized and partial epilepsy. Additionally, it can be observed that the patients with generalized and partial seizures present a higher number of early seizures, as denoted by the steeper fall of the curves at earlier years.

In the second plot to the right we only consider patients from Arm A, and stratify by epilepsy type. It is now noticeable that there is are only very few patients with generalized seizures in this category, while the great majority correspond to patients presenting partial seizures. In the third plot corresponding only to patients allocated to Arm B, we observe that in general there are fewer observations in this Arm, compared to Arm A. For this reason, in the preliminary study we are more interested in studying the patients with partial seizures in Arm A. In the fourth plot the survival curves stratified by type of drug can be observed. Except for patients in Arm A under treatments Topiramate and Oxcarbazepine, all remaining curves show a high number of early seizures.

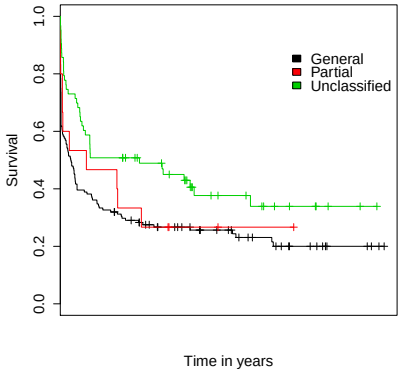
In the paper [4] by Bonnett et al, several risk factors were found to be significant for both modelling the time to treatment failure, as well as modelling the time to 12 month remission. For the time to treatment failure, which we considered for the model, the risk factors identified were: sex, treatment history (taking non-SANAD anti-epileptic drugs vs treatment naive), age, total number of seizures, electroencephalogram (EEG) results, seizure type, site of onset and the treatment (the type of drug). In our model, the site of onset and the treatment history factors



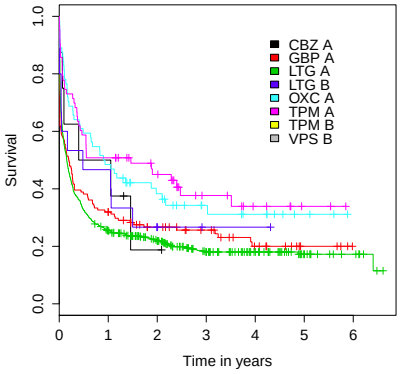
(a) Data stratified by Type of epilepsy.



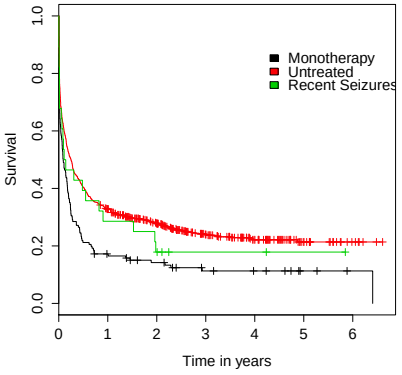
(b) Arm A, stratified by Type of epilepsy.



(c) Arm B, stratified by Type of epilepsy.



(d) Data stratified by Drug.



(e) Data stratified by Treatment history.

Figure 2.3: Kaplan-Meier survival curves for SANAD study, representing the time to first seizure since randomization, stratified by different covariates.

have not been considered, as they need to be better understood from the original data set. We see, however, in the Kaplan-Meier curves presented as the last plot in Figure 2.3, that the number of patients without previous treatment represents a great majority of the data. Indeed, in Table 2.6 we observe, for instance, that in Arm A there are 593 previously untreated patients, against 128 with monotherapy and 16 presenting recent seizures.

	Arm A	Arm B
Monotherapy	128	23
Recent seizures	16	12
Untreated	593	187

Table 2.6: Number of patients by treatment history and by Arm.

For the age factor, we consider the age at randomization of each patient. The ages range from 5 to 83 years old, with a mean of 38.3 years; for better interpretation purposes in the analysis, we center the ages at 40 years and divide in decades, so that our variable will be of the form $(age - 40)/10$. In Figure 2.4 we observe these transformed ages at randomization by drug and Arm allocation. Observe that generally patients from Arm A range from around 30 to 70 years old, while all patients from Arm B are younger and range about 20 and 30 years old. This is in turn a result from the behavior of the epilepsy types that usually affect patients in different age ranges.

In Section (4.1.5) two point-process models are proposed and fitted for this data set, considering the two arms and a comparable set of covariates to the ones used under the MESS data set. Although the MESS data set remains the object of our main interest in terms of model fit and model behavior observation, the fits corresponding to the SANAD study are done to provide a measure of comparison of interpretation.

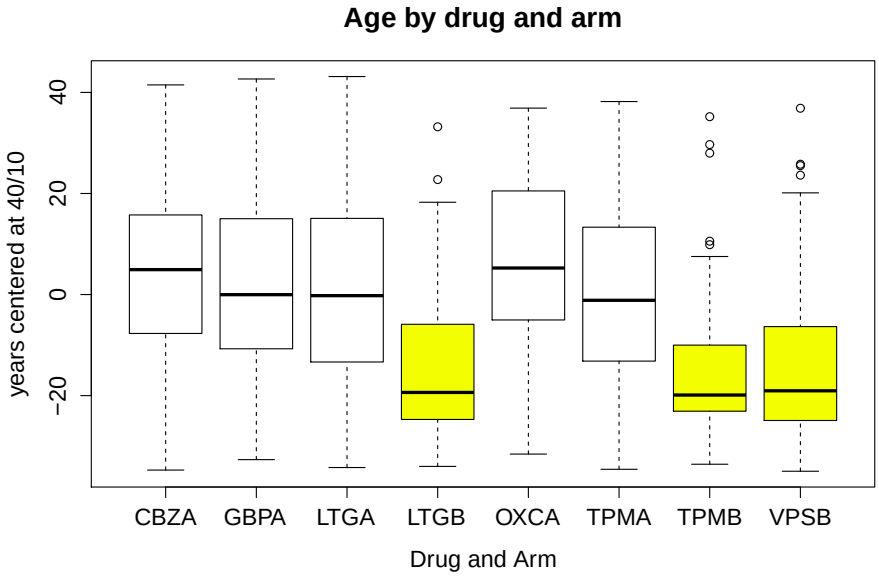


Figure 2.4: Box-plot of the ages of patients by Arm and Drug. The Box-plots are coloured in white or yellow to signal if these are plots from Arm A or Arm B respectively.

2.3 Introduction to Survival Analysis Theory

Suppose we have a population of n independent devices or units which are turned on or start living at time $t_0 = 0$, and let $T > 0$ be the time at which a random unit drawn from the population fails. The theoretical probability models used to describe the random variable T are known as "Life distributions", "Reliability Models ", or as they are more commonly known in life sciences, "Survival Models". Let $F(t)$ be the cumulative distribution function (CDF) of T, and let us denote by $f(t)$ the probability density function corresponding to $F(t)$. Then $F(t)$ has two possible interpretations:

1) $F(t)$ is the probability that a random unit drawn from the population fails by time t,

2) $F(t)$ is the proportion of units that fail by time t.

Since we are more interested in the surviving units, that is to say, those that

have not failed, we define the *Reliability function* by

$$R(t) = 1 - F(t),$$

where $R(t)$ is also written as $S(t)$ in life sciences, after its equivalent name, *Survival function*. Such function may be thought of in two ways:

1) As the probability that a random unit drawn from the population is still running by time t,

2) as the proportion of units from the population that are still operational by time t.

If n identical and independent units are operating and $F(t)$ describes the population they come from, the probability that none of the elements has failed by time t is given by $[S(t)]^n$. Consequently, the probability that at least one element fails by time t is given by

$$1 - [S(t)]^n = 1 - [1 - F(t)]^n .$$

2.3.1 Hazard Function

Given that we have a population of n independent objects, all of which have the same reliability $R(t)$, our next natural concern is to find the "rate of failure" of such objects. That is, the proportion of units that will fail by unit of time. For this, observe that the probability that an object has not failed by time t, but fails at time $\Delta t > t$, is

$$\frac{P(\text{fail at } \Delta t | \text{survive until time } t)}{\Delta t} = \frac{F(t + \Delta t) - F(t)}{S(t)\Delta t},$$

and from this, as we make $\Delta t \to 0$, we define the *instantaneous failure rate* or *hazard function*, $h(t)$, as

$$h(t) = \frac{f(t)}{S(t)}.$$

Observe that the hazard function is always non-negative, but it is not restricted to be less than one. This hazard function can be integrated to find the *Cumulative hazard function* $H(t)$ given by

$$H(t) = \int_0^t h(y)dy = -\ln S(t),$$

where "ln" denotes the natural logarithm. Let us notice that such integral must then tend to infinity as time tends to infinity, since $F(t)$ must tend to one as this happens, and we have the useful identity relating failure rates and CDFs as follows:

$$F(t) = 1 - e^{-H(t)} = 1 - e^{-\int_0^t h(y)dy}.$$

Observe, from this result, that then we can obtain all quantities from any of the expressions $f(t)$, $F(t)$, $R(t)$, $h(t)$ or $H(t)$. They all provide the same information.

The shape of the hazard function indicates how a unit of the population ages. Intuitively it indicates to how much risk is the unit subjected as time passes. Although any nonnegative function $h(t)$ whose integral $H(t)$ approaches infinity as time tends to infinity can be a hazard function, the three usual shapes of such function are the *increasing hazard function* (labeled IFR for *increasing failure rate*), the *decreasing hazard function* (labelled DFR after *decreasing failure rate*) and the *Bathtub* shaped hazard function. The IFR failure rate is used to model objects that tend to fail more often when time increases. Similarly, the DFR failure rate models cases in which the units tend to fail in the early stages. And finally, the *Bathtub* shaped hazard function is used to model the situation in which the individual fails more frequently at the beginning, then it stabilizes, and due to some degradation

process, it tends to fail more as time increases. This *Bathtub* shaped model is commonly used when dealing with living beings lives.

Since the hazard function $h(t)$ varies over time, the natural way to define the *Average Hazard Rate,* or *Average Failure Rate (AFR)* between time t_1 and time t_2 is

$$AFR(t_1, t_2) = \frac{\int_{t_1}^{t_2} h(t)dt}{t_2 - t_1} = \frac{H(t_2) - H(t_1)}{t_2 - t_1} = \frac{\ln S(t_1) - \ln S(t_2)}{t_2 - t_1};$$

and if the interval under consideration is from 0 to T, the AFR becomes

$$AFR(T) = \frac{H(T)}{T} = \frac{-\ln S(T)}{T}.$$

This quantity tells us how many units will, on average, fail in the interval of time 0 to T. This expression, however, is seldom used in practice, but it is helpful when we have an interest in the specifications of a unit, for example, in the industry environment.

In order to define another expression of interest, let us first observe that the mean of a lifetime distribution is given as

$$\mu = \int_0^\infty t f(t)dt,$$

which is known as the *Mean time to fail* ($MTTF$). This is the average time to failure for all the failure times in the population. Now, we may be interested on the average lifetime of a unit who has survived up until time T_0; this is defined as the *Mean residual time* ($ResidualMTTF$), and is given by the expression

$$ResidualMTTF(T_0) = \int_{T_0}^\infty t \frac{f(t)}{S(T_0)} dt = \frac{\int_{T_0}^\infty S(t)dt}{S(T_0)}.$$

2.3.2 Proportional Hazards and Accelerated Life Models

If we consider a set of covariates $x = (x_1, x_2, ..., x_p)^T$, then the *general proportional hazards model* is defined as

$$h(t|x_1, ..., x_p) = h_0(t)g(x_1, ..., x_p),$$

where $h_0(\cdot)$ is the *underlying hazard function*, also called the *base-line hazard function*, and $g(x_1, ..., x_p)$ represents the effect from the covariates. As a more particular model, the Cox proportional hazards model assumes that $g(x) = \exp\left(\sum_{j=1}^{p} b_j x_j\right) = \exp\left(b^{'}x\right)$, where $b = (b_1, ..., b_p)$ are the coefficients of the covariates. This means that, under Cox's proportional hazards, the model is of the form

$$h(t|x) = h_0(t) \exp\left(b^{'}x\right).$$

Another popular survival model is the accelerated failure time model, also known as the accelerated life model. It considers once more that, if b and x are considered as in the previous definition, then the model is defined as

$$h(t) = \exp(b^{'}x)h_0(t \exp(b^{'}t)),$$

where $h_0(\cdot)$ is the base-line hazard function.

When we have a data set in which all the lifetimes of the components are known, then this is called a *complete data set*. However, it often happens that, in lifetime data, it is impractical or impossible to obtain or observe the lifetimes of all objects in the observed population. We say that an observation is censored when only one bound of its lifetime is known. From this, we have a *censored data set* when such set contains one or more censored observations. In turn, there are two types of censoring: those with exact failure times or those with Readout time or

interval (or grouped) data.

Right Censoring

Three special cases of right censoring are often observed in the Survival models situations. In all three cases, it is possible to record the exact failure times of the non-censored units.

Right Censoring Type I

This kind of censoring is also called *time censoring,* for the following reason. Suppose we have n units that are put on test for a fixed planned duration of time T, at which the test is put to an end. Say that $r < n$ fail, and their failing times $t_1 \leq t_2 \leq ... \leq t_r \leq T$ are recorded. Then, at the end of the test, we have r units which failing times have been observed, and $n-r$ units that survived the test. From these $n-r$ objects we only know that their failing times are greater than T, and in here, the number of failures r is a random quantity.

Right Censoring Type II

This type of censoring, also called the *order statistic censoring*, corresponds to terminating the study upon one of the ordered failures. Once more suppose we have n units on test, and we wait until exactly r failures occur and then stop. Since r is specified in advance, such quantity is no longer random. However, now the length of the test is random and open ended, which makes this kind of censoring much more impractical than the time censoring.

Right random censoring

This occurs when individual items are withdrawn from the test at any time during the study. It is usually assumed that the ith lifetime t_i and the censoring time c_i are independent random variables. This causes that the units are not censored because they are at unusually high or low risk of failure.

2.3.3 Readout Time or Interval Data

The recollection of the precise failing times requires instruments that continuosly monitor the components. However, this may be impractical or too costly in many cases. That is a reason why the following censoring scheme is very common: suppose we have n components that are put on test at time zero. At time T_1, a "readout " takes place whereby all the components are examined and failures are removed. Let us say r_1 failures are found. Then, $n - r_1$ components go back on the test. At time T_2, after a lapse of time $T_2 - T_1$, another readout is performed and once again the new r_2 failures are removed, and so this process goes on. The last readout, at time $T_k = T$, takes place at the end of the test.

In this kind of censoring, the readout times are predetermined, but the exact times of failure are never known, which in turn makes us lose precision. An additional problem with readout data experiments is that the experiment may end before a sufficient number of failures takes place. Even if there are many failures, the data may be deficient if these failures occur in too few intervals.

Even with all these drawbacks and the inherent difficulty of analyzing this kind of data sets, this is probably the most common type of reliability data, since it is the most practical one in many situations.

Left Censoring

This happens less frequently than Right censoring, and is also a censoring in which is possible to record the exact failing times of the non-censored units. For a sample of n units, exact times of failure may occasionally be missing for the earliest failures. All that is known is that r failures took place before time T, and from then on, failure or censoring times were recorded on the remaining $n-r$ units. Left-censored data arise for failures occuring before the first readout for interval data, and we only know that the failures happened before the first inspection.

2.4 Likelihood Theory

Let $t_1, t_2, ..., t_n$ be lifetimes sampled randomly from a population of items with a lifetime distribution having a probability density function $f(t;\theta)$. This distribution has a vector $\theta = (\theta_1, ..., \theta_p)^T$ of unknown parameters associated with it, where p is the number of unknown parameters. Since the lifetimes are independent, the likelihood function, $L(t,\theta)$, is given by

$$L(t,\theta) = f(t_1, ..., t_n; \theta) = \prod_{i=1}^{n} f(t_i;\theta),$$

where $t = (t_1, ..., t_n)$. In turn, the maximum likelihood estimator is given by

$$\hat{\theta} = \max_{\theta \in \Theta} L(t;\theta).$$

In practice, it is usually easier to maximize the log-likelihood function $\log L(t;\theta)$, by solving the differential equation

$$S(t;\theta) = \left(\frac{\partial}{\partial \theta_1} \log L(t,\theta), ..., \frac{\partial}{\partial \theta_p} \log L(t,\theta) \right)^T = (0, ..., 0)^T,$$

where $S(t;\theta)$, the vector of partial derivatives of the log-likelihood function, is known as the Score function. The maximum likelihood estimator $\hat{\theta}$ has a variance-covariance matrix given by the so called Fisher Information matrix:

$$I(\theta) = E\left[S(t;\theta) S(t;\theta)^T \right].$$

This matrix has components

$$E\left[\frac{-\partial^2 \log L(t;\theta)}{\partial \theta_i \partial \theta_j} \right], \; i = 1, 2, ..., p, \quad \text{and} \quad j = 1, 2, ..., p.$$

However, we usually use the Observed Information matrix,

$$O(\hat{\theta}) = \left[\frac{-\partial^2 \log L(t;\theta)}{\partial\theta_i \partial\theta_j}\right]_{\theta=\hat{\theta}} \quad \text{for} \quad i = 1, 2, ..., p \text{ and } j = 1, 2, ..., p.$$

This point estimation is often used because of its invariance principle, meaning that if there is a one-to-one function τ, then the maximum likelihood estimator of $\tau(\theta)$ is $\tau(\hat{\theta})$ if $\hat{\theta}$ is the maximum likelihood estimator of θ.

2.4.1 Censoring

A typical assumption when constructing the likelihood function for censored data, is that the lifetimes and the censoring times are independent. If they are not independent, some special techniques are required. In the construction of the likelihood function, we need to consider what kind of information each type of censoring give us. We know that for the noncensored elements, each observation provides the information of the exact time at which the failure occurred. This provides the probability that the event happened at this time, which is approximately equal to the density function of X at the failure moment.

For a right-censored observation, all we know is that the failure time is greater than this time, hence the information is the reliability function evaluated at the on study time. Similarly, for a left-censored observation, all we know is that the event has already occurred, so the contribution to the likelihood is the cumulative distribution function evaluated at the on study time. Finally, for interval-censored data, we only know that the event occurred within the interval in question, hence we use the probability that the event time is in the interval.

In other words, the likelihood function may be constructed by putting to-

gether the component parts as

$$L(t;\theta) \propto \prod_{i \in D} f(t_i;\theta) \prod_{i \in R} S(C_r;\theta) \prod_{i \in L} [1 - S(G;\theta)] \prod_{i \in I} P\left[(a,b);\theta\right],$$

where we have n observations, $t = (t_1, ..., t_r)^T$ are the r noncensored failing times, D is the set of indexes of noncensored objects. R is the index set of the right-censored units by time C_r, L is the set of indexes of left-censored units by time G, and I is the index set of the objects with failing times censored in intervals with endpoints (a, b). The unknown parameters for this setup are denominated by $\theta = (\theta_1, ..., \theta_p)^T$.

The usual representation of data that comes from a censored experiment is (T, δ), where T and δ are random variables such that, if X is a lifetime:

$$\delta = \begin{cases} 1 \text{ if } X \text{ is observed,} \\ 0 \text{ if } X \text{ is not observed,} \end{cases} \quad \text{and} \quad T = \begin{cases} X \text{ if the lifetime is observed,} \\ C_r \text{ if the lifetime is right-censored.} \end{cases}$$

Or, in other words, $T = \min(X, C_r)$.

We may now construct the likelihood function for the Type I censoring as follows. For $\delta = 0$,

$$\begin{aligned} P(T, \delta &= 0) = P(T = C_r | \delta = 0) P(\delta = 0) = P(\delta = 0) \\ &= P(X > C_r) = S(C_r). \end{aligned}$$

Now, for $\delta = 1$,

$$\begin{aligned} P(T, \delta &= 1) = P(T = X | \delta = 1) P(\delta = 1) \\ &= P(T = X | X \leq C_r) P(X \leq C_r) \\ &= \left(\frac{f(t)}{1 - S(C_r)} \right) (1 - S(C_r)) = f(t). \end{aligned}$$

These expressions can be combined into the single form

$$P(t, \delta) = [f(t)]^{\delta}[S(t)]^{1-\delta}.$$

If we have a random sample of pairs (T_i, δ_i), $i = 1, 2, ..., n$, then the likelihood function is

$$L(t_1, ..., t_n, \delta) = \prod_{i=1}^{n} P(t_i, \delta_i) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i},$$

where $\sum_{i=1}^{n} \delta_i = r$. Since we can write $f(t_i) = h(t_i)S(t_i)$, we can write the likelihood function as

$$L(t_1, ..., t_n, \delta) = \prod_{i=1}^{n} [h(t_i)]^{\delta_i} \exp[-H(t_i)].$$

In a very similar way, we find that the likelihood function corresponding to the observation of exact and left-censored data, is given by

$$L(t_1, ..., t_n, \delta) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i}[1 - S(t_i)]^{1-\delta_i}.$$

And finally, for data coming from an interval-censoring, if we censored all items with failing times falling outside of the intervals (a_i, b_i) for $i = 1, ..., n$, then the likelihood function for this kind of censoring is

$$L(t_1, ..., t_n, \delta) = \prod_{i=1}^{n} [f(t_i)]^{\delta_i} P[(a_i, b_i)]^{1-\delta_i}.$$

2.5 Parametric Lifetime Models

Three types of parameters are used in the distributions described later on: location, scale and shape. *Location* (or *shift*) parameters are used to shift the distribution to the left or right along the time axis. If c_1 and c_2 are two values of a location parameter of a lifetime distribution with survival function $S(t; c)$, then there exists

a real constant α such that $S(t; c_1) = S(t + \alpha; c_2)$. A familiar example of a location parameter is the mean of the normal distribution.

Scale parameters are used to expand or contract the time axis by a factor of α. If λ_1 and λ_2 are two values of a scale parameter for a lifetime distribution with risk function $S(t; \lambda)$, then there exists a real constant α such that $S(\alpha t; \lambda_1) = S(t; \lambda_2)$. A familiar example of a scale parameter is λ in the exponential distribution. The probability density function always has the same shape, and the units on the time axis are determined by λ.

Shape parameters are appropriately named since they affect the shape of the probability density function. Shape parameter values may determine whether a distribution belongs to a particular distribution class such as IFR (Increasing Failure Rate) or DFR (Decreasing Failure Rate). A familiar example of a shape parameter is κ in the gamma distribution.

2.5.1 Exponential Distribution

The exponential distribution is presented first due to its simplicity. This distribution has a single positive scale parameter λ, often called the failure rate, and its probability density function and cumulative probability distribution are, respectively,

$$\begin{aligned} f(t; \lambda) &= \lambda e^{-\lambda t}, \text{ for } t > 0 \text{ and } \lambda > 0, \\ F(t; \lambda) &= 1 - e^{-\lambda t}. \end{aligned}$$

Since $R(t; \lambda) = 1 - F(t; \lambda) = e^{-\lambda t}$, the failure rate function or hazard function for this distribution is

$$h(t; \lambda) = \frac{f(t; \lambda)}{R(t; \lambda)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

This result shows that the exponential failure rate function reduces to the value λ for all times. This is a characteristic property of the exponential distribution. The

only distribution with a constant hazard function is the exponential, since

$$F(t;\lambda) = 1 - e^{\int_0^t h(y)dy},$$

if $h(y) = \lambda$, a constant, the integral becomes

$$H(t;\lambda) = \int_0^t \lambda dt = \lambda t,$$

then we have $F(t;\lambda) = 1 - e^{-\lambda t}$, or the exponential CDF.

The exponential distribution has often been used to model the lifetime of electronic components and is appropriate when a used component that has not failed is statistically as good as a new component. A brand new unit has a expected lifetime until failure described by the $MTTF$:

$$MTTF = \int_0^\infty R(t)dt = \int_0^\infty e^{-\lambda t}dt = \frac{1}{\lambda}.$$

We interpret this result as follows: the $MTTF$ for a population with a constant failure rate λ is a reciprocal of the failure rate or $1/\lambda$.

Although $1/\lambda$ is the average time to failure, it is not the same as the time when half the population would have failed. This median time to failure, T_{50}, gives us

$$F(T_{50}) = 0.5 = 1 - e^{\lambda T_{50}},$$

taking natural logarithms and solving for T_{50}, we have

$$T_{50} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda}.$$

If we feel that a unit under test has no significant wear-out mechanisms, at least for its intended application life, and either we do not expect many early defect failures or we intend to separate these out and treat them separately, then

the exponential is a good initial choice for a life distribution model.

Estimation of the Exponential Failure rate λ

When data come from an exponential distribution, there is only one parameter, λ, to estimate. The best estimate for complete or censored samples is

$$\hat{\lambda} = \frac{\text{number of failures}}{\text{total unit test hours}}.$$

The denominator is the sum of all the operation hours of every unit on the test, including both failures and those that have completed the test without failing. For a complete sample, this expression reduces to the reciprocal of the sample mean. Thus, we have $\hat{\lambda} = 1/(\text{sample mean time to failure})$, just as we had $\lambda = 1/MTTF$.

The maximum likelihood estimators of λ, considering the Type I right-censored experiment, will be described as follows. Remember that the likelihood equation consists of two parts: (1) the product of the r PDFs for each failure time, and (2) the survival probability at the end of test raised to the power of the number surviving, $n - r$:

$$\begin{aligned} L(t;\lambda) &\propto \left[\prod_{i=1}^{r} f(t_i)\right] [1 - F(T;\lambda)]^{n-r} \\ &= \lambda^r \exp\left(-\lambda \sum_{i=1}^{r} t_i\right) \left(e^{-\lambda T}\right)^{n-r}, \end{aligned}$$

Then, the log-likelihood is given by

$$l(t;\lambda) = -(n-r)\lambda T - \lambda \sum_{i=1}^{r} t_i + r \ln \lambda.$$

In turn, the Score function is then found and equaled to zero:

$$S(t;\lambda) = \frac{\delta}{\delta\lambda} l(t;\lambda) = \frac{r}{\lambda} - (n-r)T - \sum_{i=1}^{r} t_i = 0$$

and by solving such equation we obtain

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_i + (n-r)T},$$

where T is the pre-specified end of test time and $t_1, t_2, ..., t_r$ are the exact failure times of the r units that fail before the test ends.

If the test is censored Type II (ends at rth failure time t_r), the same rule yields

$$\hat{\lambda} = \frac{r}{\sum_{i=1}^{r} t_i + (n-r)t_r}.$$

When we have readout data, we can no longer exactly calculate the denominator in order to estimate λ. In this case, graphical methods can be used, which also apply as an alternative approach, when exact times are available.

Exponential Distribution Closure Property

A system model where n components operate independently and the system fails with the first component failure is called a series model. It can be shown that the system failure rate, or hazard function $h_s(t)$, is the sum of the n component hazard functions $h_1(t), h_2(t), ..., h_n(t)$. When the components have exponential lifetimes with parameters $\lambda_1, \lambda_2, ..., \lambda_n$, then the system has a constant failure rate equal to

$$\lambda_s = \sum_{i=1}^{n} \lambda_i$$

This result establishes the exponential closure property since a constant hazard function implies an exponential distribution. If the components are all the same, each having hazard function λ, then the system has a hazard function $n\lambda$ and an $MTTF$ of $1/n\lambda$.

2.5.2 Weibull Distribution

The exponential distribution is limited in applicability because of the memoryless property. The assumption that a lifetime has a constant failure rate is often too restrictive or inappropriate. Mechanical items typically degrade over time and hence are more likely to follow a distribution with a strictly increasing hazard function. The Weibull distribution is a generalization of the Exponential distribution that is appropriate for modeling lifetimes having a constant, strictly increasing, and strictly decreasing hazard functions.

The probability density function (pdf) $f(t;\kappa,\lambda)$, the risk function $R(t)$, the hazard function $h(t)$ and the cumulative hazard function $H(t)$ corresponding to the Weibull distribution are as follows:

$$f(t;\kappa,\lambda) = \kappa\lambda^{\kappa}t^{\kappa-1}e^{-(\lambda t)^{\kappa}}, \quad S(t) = e^{-(\lambda t)^{\kappa}},$$
$$h(t) = \kappa\lambda^{\kappa}t^{\kappa-1}, \qquad H(t) = (\lambda t)^{\kappa},$$

for all $t \geq 0$, where $\lambda > 0$ and $\kappa > 0$ are the scale and shape parameters of the distribution, respectively.

Observe that, when $\kappa = 1$, the hazard function is a constant. Thus, the Weibull reduces to an exponential with hazard function λ.

The Weibull CDF equation has four quantities that may be known, assumed, or estimated from data. These include the cumulative fraction failed $F(t;\kappa,\lambda)$, the time t , the shape parameter κ, and the scale parameter λ. If any of these three are known, the fourth can be calculated. Now, using maximum likelihood estimation, observe that, if $t_1, ..., t_n$ are the failure times, $c_1, c_2, ..., c_n$ are the censoring times, and $x_i = \min\{t_i, c_i\}$ for $i = 1, ..., n$, and U is the set of uncensored units, then the

log-likelihood function is

$$\begin{aligned} l(t;\kappa,\lambda) &= \log L(t;\kappa,\lambda) = \sum_{i\in U} \log h(x_i;\kappa,\lambda) - \sum_{i=1}^{n} H(x_i;\kappa,\lambda) \\ &= \sum_{i\in U} (\log\kappa + \kappa\log\lambda + (\kappa-1)\log x_i) - \sum_{i=1}^{n} (\lambda x_i)^{\kappa} \\ &= r\log\kappa + \kappa r\log\lambda + (\kappa-1)\sum_{i\in U}\log x_i - \lambda^{\kappa}\sum_{i=1}^{n} x_i^{\kappa}, \end{aligned}$$

and the 2×1 Score vector has elements

$$Sc_1(\lambda,\kappa) = \frac{\partial \log L(t;\lambda,\kappa)}{\partial\lambda} = \frac{\kappa r}{\lambda} - \kappa\lambda^{\kappa-1}\sum_{i=1}^{n} x_i^{\kappa}$$

and

$$Sc_2(\lambda,\kappa) = \frac{\partial \log L(t;\lambda,\kappa)}{\partial\kappa} = \frac{r}{\kappa} + r\log\lambda + \sum_{i\in U}\log x_i - \sum_{i=1}^{n}(\lambda x_i)^{\kappa}\log(\lambda x_i).$$

When this are set equal to zero, the simultaneous equations have no closed-form solution for $\hat{\lambda}$ and $\hat{\kappa}$. To avoid solving a 2×2 set of nonlinear equations, the fortunate event is that the first equation can be solved for λ in terms of κ as follows:

$$\lambda = \left(\frac{r}{\sum\limits_{i=1}^{n} x_i^{\kappa}}\right)^{1/\kappa}.$$

Using this expression, $Sc_2(\lambda,\kappa)$ now does not depend on λ, and we obtain a single expression with κ as the only unknown. Applying some algebra, this equation reduces to

$$g(\kappa) = \frac{r}{\kappa} + \sum_{i\in U}\log x_i - \frac{r\sum_{i=1}^{n} x_i^{\kappa}\log x_i}{\sum_{i=1}^{n} x_i^{\kappa}} = 0,$$

which must be solved iteratively. One technique that can be used here is the Newton-

Raphson procedure, which uses

$$\kappa_{i+1} = \kappa_i - \frac{g(\kappa_i)}{g'(\kappa_i)},$$

where κ_0 is the initial estimator. The iterative procedure can be repeated until the desired accuracy for κ is achieved.

The strength of the Weibull lies in its flexible shape as a model for many different kinds of data. Since for $0 < \kappa < 1$,

$$\lim_{t\to 0} f(t;\lambda,\kappa) = \infty \text{ and } \lim_{t\to\infty} f(t;\lambda,\kappa) = 0,$$

the hazard function behaves the same way, making this type of Weibull a useful model for an early failure mechanism. For $\kappa = 1$ the Weibull reduces to a standard exponential with constant failure rate $1/\lambda$. For $\kappa > 1$, the PDF starts at zero and increases to a peak at $\lambda[1-(1/\kappa)]^{1/\kappa}$, after which it decreases towards zero as time increases. The shape is skewed to the right when k is large. The hazard function also starts at zero but the increases monotonically throughout life. This type of Weibull is a useful model for wear-out failure mechanisms.

The Weibull also has a closure or reproductive property, similar to the exponential. If a system is composed of n parts, each having an independent Weibull distribution with the same shape parameter κ, but not necessarily the same location parameter, and the system fails when the first component fails (a series system), then the time to the first system failure also follows a Weibull distribution. If the location parameters are $\lambda_1, ..., \lambda_n$ and the shape parameter is κ, the system's location parameter is given by

$$\lambda_s = \left(\sum_{i=1}^{n} \frac{1}{\lambda_i^{\kappa}}\right)^{1/\kappa}.$$

2.5.3 Gamma Distribution

The Gamma distribution is a second important generalization of the exponential distribution. The probability density function for the gamma distribution is

$$f(t;\kappa,\lambda) = \frac{\lambda}{\Gamma(\kappa)}(\lambda t)^{\kappa-1}e^{-\lambda t}, \quad t \geq 0,$$

where λ and κ are positive scale and shape parameters, respectively. When $\kappa = 1$, the gamma distribution is equivalent to the exponential distribution. Usually it is difficult to differentiate between the Weibull and gamma distributions based on plots of their probability density function, since the shape of their plots are similar. The differences between these two distributions become apparent when their hazard functions are compared. The cumulative distribution function of a gamma random variable is

$$F(t) = \frac{1}{\Gamma(\kappa)}\int_0^{\lambda t} x^{\kappa-1}e^{-x}dx,$$

such distribution is called the incomplete gamma function. The Risk function is $R(t) = 1 - F(t)$, and as before, the hazard function is $h(t) = f(t)/R(t)$ with cumulative hazard function $H(t) = -\log R(t)$. The hazard function can be shown to be monotone increasing for $\kappa > 1$, with $h(0) = 0$, and

$$\lim_{t\to\infty} h(t) = \lambda.$$

For $0 < \kappa < 1$, $h(t)$ is monotone decreasing, with

$$\lim_{t\to\infty} h(t) = \lambda \text{ and } \lim_{t\to 0} h(t) = \infty.$$

Notice that $\lim_{t\to\infty} h(t) = \lambda$ for all values of κ, indicating that a lifetime with a gamma distribution will have an exponential tail. Thus, if an item survives far enough into the right-hand tail of the probability density function, the distribution

of the remaining time to failure is approximately exponentially distributed by the memoryless property.

Chapter 3

Previous Models for Count Data

In this chapter we present a broad introduction to previous models proposed for the MESS epilepsy data set. An important approach to medical recurrent event problems lies in the use of survival model analysis. In a large number of cases, due to familiarity and convenience, researchers tend to prefer choosing Cox's proportional hazards model as the underlying model behind such problems. However, more recent work has suggested that accelerated life models might provide a better fit in some cases. This poses the question of what is the impact of survival regression model misspecification on the parameter estimates and the conclusions they can lead to.

Indeed, the paper by Kwong and Hutton[30] addresses this problem, and mentions that it is not only important to consider the effects of model misspecification, but that it should also be noted that the choice of the baseline family distribution is often ignored. When working with epilepsy data, several important questions arise, such as which factors influence the recurrence of seizures, which base-line distributions best fit this disease, which assumptions must be made about the model, and how robust the results and interpretations are under base-line and survival regression model misspecification.

3.1 Some Historical Background

In 1997, Hutton and Solomon[25] studied the robustness of regression estimates under misspecification of the survival models. In other words, assuming a regression analysis on survival data, they were interested in the changes that the regression parameters would experience if the estimation were carried out under a proportional hazards model assumption, when the true underlying model was an accelerated life model, and vice versa. For this purpose, the main interest lies in using the orthogonality of parameters to study the effects of model misspescification. By orthogonalizing the parameters of interest from the nuisance parameters, they were able to obtain an expression of the regression parameters relating their value under the true model, and the value under the assumed model.

Using the same form of Cox & Reid's[12] mixture model, the model proposed in Hutton & Solomon[25] is a mixture of an accelerated life (AL) model, and a proportional hazards (PH) model:

$$m(t) = \left\{ g_0(t)\, e^{\beta' z} G_0(t)^{\exp(\beta' z)-1} \right\}^{\psi} \left\{ f_0\left(te^{\gamma' z}\right) e^{\gamma' z} \right\}^{1-\psi},$$

where the survivor functions $F_0\left(te^{\gamma' z}\right)$ and $G_0(t)^{\exp(\beta' z)}$ correspond to the accelerated life model and the proportional hazards model respectively; F_0 and G_0 are the base-line survivor functions with corresponding densities f_0 and g_0 . Here z is the vector of fixed covariates, and β and γ are the associated vectors of unknown regression parameters for the proportional hazards and accelerated life families of models. The interest lies in finding expressions for the parameters of interest, β and γ, while considering ψ as a nuisance parameter.

Now let us consider the following definition:

If $\theta = (\theta_1, \theta_2)$ where θ_1 and θ_2 partition θ, and $i_{\theta_i \theta_j}$ corresponds to the ij^{th}

entry of the information matrix, then we define θ_1 to be orthogonal to θ_2 if

$$i_{\theta_s\theta_t} = \frac{1}{n}E\left[\frac{\partial l(\theta;y)}{\partial\theta_s}\frac{\partial l(\theta;y)}{\partial\theta_t};\theta\right] = \frac{1}{n}E\left[-\frac{\partial^2 l(\theta;y)}{\partial\theta_s\partial\theta_t};\theta\right] = 0$$

for all $\theta_s \in \theta_1$ and for all $\theta_t \in \theta_2$.

We have from the Cox & Reid (1987) paper, shown in the Appendix under "Parameter orthogonalization", two parameters (ψ, ϕ) are orthogonal if the following equation holds

$$\sum_{r=1}^{q} i^*_{\phi_r\phi_s}\frac{d\phi_r}{d\psi} = -i^*_{\psi\phi_s},$$

where $i^*_{\psi\phi_s} = E\left[\frac{d^2 l^*(\psi,\phi)}{d\psi d\phi_s}\right]$, $i^*_{\phi_r\phi_s} = E\left[\frac{d^2 l^*(\psi,\phi)}{d\phi_r d\phi_s}\right]$ and l^* is the log-likelihood function corresponding to ψ and ϕ. These are also called orthogonalizing equations of the parameters ϕ_r and ϕ_s. Among several other convenient properties of two orthogonal parameters ψ and λ, if $\hat{\psi}$ and $\hat{\lambda}$ are the maximum likelihood estimators of ψ and λ respectively, then they are asymptotically independent.

In the case of this problem, the exact expressions of this orthogonalizing equations are not obtainable, hence the mixture function $m(t)$ is normalized and equated to the likelihood function

$$L(\psi,\beta,\gamma;\alpha,t,z) = \frac{m(t)}{\int_0^\infty m(t)dt};$$

where α, which is assumed as known, are the parameters of the base-line survival distributions.

While the formal computations are shown in the Appendix, it is worth mentioning that in order to approximate $\int_0^\infty m(t)dt$, a Taylor series is used about $(\beta,\gamma,\psi) = 0$. Observe that this requires the assumption that β and γ are small.

After several approximation steps, it was found that, if the true model is the accelerated life model, which implies that $\psi = 0$, but the assumed model is a pro-

portional hazards model, then the orthogonalizing equations result in the expression

$$\beta_s = -\gamma_s \left(1 + \frac{c_4}{c_2}\right),$$

where

$$c_2 = E_{f_0}[\log G_0(T)] \quad \text{and} \quad c_4 = E_{f_0}\left[T\dot{y}(T)\log G_0(T)\right],$$

where T is the time to occurrence random variable, and $\dot{y}(T)$ denotes the first derivative of $y(t)$ with respect to t and evaluated at T.

The expression for the regression parameters are obtained in a similar manner when the true model is a proportional model, and the accelerated life model is assumed, which corresponds to the case when $\psi = 1$. The equation in question is of the form

$$\gamma_s = -\frac{\beta_s d_6}{d_4},$$

where

$$d_4 = E_{g_0}\left[T\dot{y}(T)\log\left\{g_0(T)/f_0(T)\right\}\right] \quad \text{and} \quad d_6 = E_{g_0}\left[T\dot{y}(T)\left\{1 + \log G_0(T)\right\}\right].$$

Observe that the relationship between the regression parameters under model misspecification, in both cases, is proportional to the first degree. It can be shown that the expressions can be simplified to

$$\frac{\beta_s}{\beta_1} = \frac{\gamma_s}{\gamma_1} \text{ for } s = 2, 3, ..., 1,$$

which means that the ratios of regression coefficients are consistent.

Important results from Hutton & Solomon seem to indicate that, under the assumption of small regression coefficients, these coefficients are consistent when facing model misspecification. This result is then reconsidered in the paper by Kwong & Hutton[30] six years later, with the purpose of investigating the robustness

of such parameter estimates under model misspecification, also assuming in this work that the covariate coefficients are small.

In order to investigate the properties of misspecification in a practical manner this paper addresses two applications: monotherapy for epilepsy and cerebral palsy studies. For each one of them, four distributions are considered for the base-lines: the gamma, log-normal, log-logistic and Weibull distributions. Both proportional hazards and accelerated life models are adjusted and studied.

Although the cerebral palsy data produced results that were consistent with Hutton and Solomon's findings, the epilepsy data analysis presented a more complex case. The epilepsy study consisted of 1183 patients from five different trials, in which each person was randomly assigned either the drug carbamazepine (CBZ) or valproate (VPS). The outcomes were the times to the first seizure after the randomization, and the variables of interest were either categorical (trial, drug and seizure type) or continuous (age, drug and age interaction, and the number of seizures six months before randomization). Such variables of interest were selected by comparing the values of $-2\log(\hat{L})$ for all linear combinations of explanatory variables in Cox proportional hazards models, where L is the maximum likelihood function. Although the Drug variable was not significant in this case, it was considered for the analysis as it was of special medical interest. From the 1183 patients in the study, 836 (71%) showed a first seizure post-randomization, and the remaining 347 (29%) patients had censored observations.

Since neurologists believed that CBZ is a more effective drug for patients with partial seizures, and VPS is better for cases with general seizures, the analysis was directed to compare the survival between the groups combining CBZ and VPS drugs with generalized and partial seizures. In all four models considered, it was found that patients with generalized epilepsy had twice as long until their first seizure post-randomization than those patients with partial epilepsy.

The proportional hazards model was shown to be a bad model for this study,

as the proportional hazards constraint on the Weibull base-line distribution failed to reflect the difference in hazard rates for generalized *versus* partial epilepsy. If proportionality constraint were true, the hazards ratio between two groups should be constant with time, but in the life-table obtained, the estimates of the hazard function of the generalized epilepsy group and the estimates corresponding to the partial seizures group did not appear to be this way. The smaller AICs in the gamma, log-normal and log-logistic models indicated that accelerated life models fitted the data better than proportional hazards models. When considering the 95th, 90th and 85th percentiles of the survival, the Weibull model is shown to greatly underestimate the survival.

Finally, for this data, it was observed with a Kaplan-Meier approximation, that the failure distribution was highly skewed, since there was a large number of early seizures. As a result, after considering small covariate coefficients, there is evidence to suggest that the regression estimates under misspecification of the models were not proportional to first-degree. Thus for this study, the results from Hutton & Solomon (1997) do not hold. Indeed, observe that, if h_{AL} is the accelerated life hazard function, then

$$h_{AL}(t) = e^{\beta' z} h_0(e^{\beta' z} t) \simeq e^{\beta' z} h_0 \left[(1 + \beta' z)t\right]$$

when approximating $e^{\beta' z}$ with its first Taylor series term. Furthermore, when approximating h_0 by its Taylor expansion about t, we obtain

$$\begin{aligned} h_{AL}(t) &\simeq e^{\beta' z} \left\{ h_0(t) + \beta' z t h_0'(t) \right\} \\ &= e^{\beta' z} h_0(t) + \beta' z t e^{\beta' z} h_0'(t). \end{aligned} \tag{3.1}$$

Observe that the first term of (3.1) is an approximation to a Cox proportional hazard function, and thus Hutton & Solomon's result will hold only if the second term of the equation is zero. If the hazard function is decreasing (increasing) rapidly, $h_0'(t)$ will

be positive (negative) and large, and cannot be ignored, then the function $h_{AL}(t)$ is not approximately a Cox's proportional hazard, and consequentially, Hutton & Solomon's consistency of the parameters does not hold. This result also proposes that estimates of regression parameters corresponding to survival models with late survival times are robust under misspecification, but if the model was such that the survival distribution is strongly right-skewed, with a short median survival, the parameters are not robust.

Precisely in this example of epilepsy data, the survival Kaplan-Meier plots showed that the hazard decreases rapidly due to the abundance of early seizures. Indeed, for this case the regression parameters do not appear to be consistent for proportional hazards and accelerated life models.

After calculating the maximum likelihood parameter estimates for the four survival models, it was found that the effects of partial epilepsy, age, the natural logarithm of the number of seizures within 6 months before randomization, and the drug-age interaction, tend to reduce the time to a first seizure post-randomization, while VPS and higher age at randomization increase the time to first seizure.

It is worth mentioning that, although there was a small hint that for the drug effect, VPS is better than CBZ, this effect should not be considered on its own since the interaction age-drug was very significant. The estimated effect of age is much more marked with CBZ than with VPS in all models. For young people, the age below which is best to use VPS varies from 23 years (for gamma and Weibull models) to 17 years old (log-logistic model).

Taking this result by Kwong and Hutton, the same year Hutton & Monaghan[24] investigated the effect of misspecifying fully parametric proportional hazards and accelerated life models, taking special interest in observing the impact on the covariate, shape and scale parameters under such models. An important contribution from this paper is the fact that they do not assume that the regression coefficients are "small".

Three categories of misspecification are considered for this paper: when the base-line distribution is false, but the covariate effects are well specified, when the covariate effects are false but the right base-line distribution is considered, and finally when both covariate effects and base-line distribution are misspecified. Additionally, the three distributions considered here for the base-line distribution are: log-normal, log-logistic and Weibull distributions. In this paper, they study the asymptotical distribution of the maximum likelihood (ML) estimator, under the misspecification assumption. Once again, further details can be observed in the Appendix, but the main idea develops from a previous result by (Cox, 1972)[12]; it states that under misspecification, the ML estimator $\hat{\beta}$ of the covariate, shape and location parameters β, are asymptotically distributed as $N(\beta_\alpha, \frac{1}{n}C(\beta_\alpha))$, where $C(\beta_\alpha) = A^{-1}BA^{-1}$, A is the Fisher information matrix and $B(\alpha)_{jk} = E_f\left[\frac{\partial \log L}{\partial \beta_j}\frac{\partial \log L}{\partial \beta_k}\right]$, with L the likelihood function. Additionally, β_α is the solution of a set of expected score equations.

After all different types of misspecifications are considered and the parameters are estimated, a large series of detailed results are stated for every base-line distribution considered. Among the most important results, it was found here that the shape and regression parameters are biased when proportional hazards models are fitted to accelerated life models. Furthermore, the accelerated life model is more robust to misspecification because of its log-linear form. Consistently with Hutton & Solomon[25], it was found that the effect of misspecification decreases as the centre of density of survival times move away from zero, which means that there are not too many early failures.

3.1.1 Epilepsy seizure recurrence models

From the previous background, we know that the choice of a survival model for medical recurrent events is important, as well as the choice of the parametric base-line distribution. When modelling for epilepsy seizure recurrence, the large amount of information can present a challenge, as it has been demonstrated by previous

models that fail to utilize all of the available information.

Medical epilepsy studies tend to have information of the patient's epilepsy history, previous to being subjected to a treatment. This provides an additional source of information, that in the past has been considered merely as a covariate for the model. Later models have considered to approach the epileptic seizure counts previous to treatment, jointly with the seizure observations after treatment, as time dependent variables.

Previous approaches have been made to medical problems where the main interest of study is to model phenomena with recurrent events. Some useful approaches were considered:

- Hougaard et. al [23] used an overdispersed Poisson distribution to model epileptic seizure counts, since in this case the mean is usually smaller than the variance. They used the power variance family as the mixing distribution, but the models did not adjust for explanatory variables.

- Marshall & Olkin [34] proposed that, if we considered Y_1 as a count variable before randomization, and Y_2 as the count variable post-randomization, then they could be jointly model (Y_1, Y_2) as

$$f(y_1, y_2|z_1, z_2) = \int_0^\infty f_{Y_1}(y_1|z_1, \nu) f_{Y_2}(y_2|z_2, \nu) g(\nu) d\nu,$$

  where f is the probability function of a bivariate negative binomial, f_1 is the probability function of a Poisson distribution with parameter $\mu_1\nu$ and dependent on the covariate vector z_1, f_2 is a Poisson distribution with parameter $\mu_2\nu$, dependent on a covariate vector z_2, and g is a gamma distribution with parameters α and $1/\alpha$.

- Cameron & Trivedi [5] considered a bivariate Poisson distribution, and take the heterogeneity between two event times as correlated, not necessarily identical.

- Cook & Lawless [8], reviewed the analysis for repeated data with intensity functions, but did not discuss an analysis of the joint variables (seizure counts, gap times between seizures).

- Cook & Wei [9] described a bivariate conditional semiparametric approach for count data, mainly, for the variables event counts and event times.

One paper of particular interest to us is Cowling, Hutton & Shaw[11]. The main aims of this study were to present a contrast between the treatment effects, and study the interaction between treatments and the covariates. In this work, the covariates are age, sex, trial indicator, and type of epilepsy. Although the standard survival analysis method uses the seizure counts pre-randomization as a fixed covariate, the model proposed in this paper treats these counts and the post-randomization counts as jointly distributed. This model is chosen in order to account for the variation within individuals.

In this paper, the meta-analysis of longitudinal survival data from five trials of two epilepsy treatments is considered. In such treatments, patients were asked to recall seizures from at least six months prior a controlled randomization of the drugs to be administered. In this randomization, the patient would be allocated either the drug carbamazepine (CBZ) or the drug valproate (VPS), which in this case, are the main focus of the investigation. The data set contains the number of seizures suffered per patient during the six months prior to randomization, drug indicator, age at randomization, sex, type of epilepsy and the individual's times to first seizure post-randomization.

It is desired to model, for each individual, an event count over a fixed initial period, jointly with a survival time following a possible change in event rate.

3.2 MESS Study First Proposed Analyses

According to Marson et al.[35], almost all patients were prescribed CBZ or VPZ, 92% in the immediate group, and 83% of those receiving AEDs in the deferred group. This paper introduces and describes the MESS study and endeavours to supply a prognosis for patients within the study. It is mentioned that the study was conducted as a randomized, unmasked and pragmatic trial, but later on it was found that the unmasking does not appear to generate a bias for the study, judging by the comparison between groups on first seizure recurrence, such recurrence did not appear to be different between them.

Although the initial intention was to recruit 3000 patients for the study in order to perform a split-half cross-validation analysis, only 1443 patients were included and accepted to randomization. As a consequence, this paper focuses on performing the analysis with the population provided, and suggests that performing a cross-validation procedure should not carried out, since the power for generating a validation sub-population diminishes greatly. However, it is noticeable that in a subsequent paper by Kim et al (2006)[27], this same MESS study is re-addressed with a cross-validation analysis.

The analysis of the first time to event is performed with a log-rank test, or Cox's proportional hazards model when adjusting single versus multiple seizures at entry. All analyses were by intention to treat, and the results are reported as absolute differences in proportions or hazard ratios with 95% confidence intervals.

For the times to first and second seizures post-randomization, the difference between treatment groups (immediate versus deferred) is highly significant, while for the time to the fifth seizure, there is no perceptible difference between groups. As for the remission period, the actuarial estimate of achieving a 2 year remission by the 8th year is very high (over 94% for both treatment groups and single versus multiple seizures). Although the immediate treatment scheme increased the times

	Immediate (%)		Deferred (%)		Difference
	Single	Multiple	Single	Multiple	(95% CI)
Time to first seizure					
6 months	18	26	26	44	12 (7.4, 16.5)
2 years	32	43	39	61	11 (6.2, 16.7)
5 years	42	57	51	69	10 (4.5, 16)
8 years	46	60	52	72	9 (2.6, 15.3)
Time to second seizure					
6 months	14		19		5 (1.1, 8.9)
2 years	24		32		8 (3.6, 13.3)
5 years	34		40		6 (0.9, 12)
8 years	38		44		5 (−1.4, 12)
Time to fifth seizure					
6 months	6		7		1 (−1.3, 3.8)
2 years	12		15		3 (−1.1, 6.2)
5 years	19		22		3 (−1.6, 7.6)
8 years	26		25		−1 (−9.3, 7.8)
2-year remission					
2 years	69	57	61	39	12 (6.3, 17.4)
5 years	92	91	92	87	2 (−1.2, 6.1)
8 years	95	94	96	95	1 (−2.5, 3.9)

Table 3.1: MESS study seizure counts, table from Kim et al[27]. The proportion of patients allocated to deferred or immediate treatment are shown here tabulated by time to seizure and seizure characterization (either single or multiple seizures).

to first and second seizures post-randomization, and decreased the time to 2 year remission, there is no significant difference between the immediate and the deferred treatments for the time to 5 year remission (Table 3.1).

As a general result, in Marson et al[35], it is mentioned that the benefits from immediate treatment do not impact on the long-term conditions of the patients, but findings from a subpopulation of 441 UK adult patients show that they do come at some cost for the patient. Indeed, from the 527 eligible UK patients, 441 of them returned a quality of life (QOL) question sheet at 100 days from randomization. This data was later on used to analyse the changes between baseline and 2-year follow-up for anxiety, depression and mastery (an individual's perception of control over his life). The immediate treatment patients reported more adverse events that were

related to the treatment, such as depression, anxiety, gastrointestinal symptoms, among others.

One year later, Kim et al[27] considered the effects of the two main anti-epileptic drugs, carbamazepine and valproate, with the objective of studying the roles of treatment policies and patient characteristics in the subsequent risk of seizure recurrence. They took a cross-validation sample using only a subset of the total population to assess the model, while the remaining subset was then used to validate such model. From the 1443 patients, 885 of them were assigned as a test sample, 535 patients were used for validation, and the remaining subjects were discarded from the study, as they had incomplete data. The population was then split into three categories (low, medium and high risk), and stratified by randomized treatment policy.

The test sample was used to formulate the prognostic model, which considered the time from randomization to the first seizure hence, based on a Cox regression and stratified by treatment allocation. Both backward and forward stepwise regressions were used to find the predictive value of the baseline covariates of interest. Each stepwise regression gave place to a different model, given that a covariate was included in the model if they were significant in univariate analysis at $p \leq 0.05$ and it was excluded if $p \geq 0.1$. Since the number of patients with several types of neurological disorder was low, all conditions (neurological deficit or impairment, delayed development and learning disability) were represented under one variable only, which is termed the Neurological disorder, and is denoted as NDL. In the same way, the variable called EEG contained three Electro-encephalogram (EEG) measurements (specific local, generalized epileptiform and slow wave abnormality).

The resulting models were Model 1 and Model 2, where the covariates considered resulted from the backward and forward stepwise elimination respectively. Stratification by treatment was used on both models. The hazard ratios for both models are shown in Table 3.2.

Variables	Model 1		Model 2	
	Hazards ratios	p-value	Hazards ratios	p-value
Neurological disorder	$1.35(1.07, 1.72)$	$p = 0.013$	$1.36(1.07, 1.73)$	$p = 0.012$
Abnormal EEG	$1.54(1.27, 1.86)$	$p < 0.0001$	$1.53(1.26, 1.86)$	$p < 0.0001$
Log(Number of seizures) pre-randomization	$1.56(1.42, 1.72)$	$p < 0.0001$	$1.61(1.46, 1.77)$	$p < 0.0001$
Years between seizure and randomization			$0.92(0.84, 1.02)$	$p = 0.1$
First degree relative with epilepsy			$1.27(0.96, 1.7)$	$p = 0.1$

Table 3.2: Resulting hazard ratios with their respective 95% confidence intervals and p-values for the forward (Model 1) and backward (Model 2) stepwise elimination, taken from Kim et al. [27].

The hazard ratios can be interpreted as the relative change in risk for a unit increase in the prognostic factor. Observe that the second model contains the variables of the first model and it has two additional variables. However, the identification of two large outliers in the data resulted in the exclusion of the variable of years between recent seizure and randomization. Furthermore, since the two variables added in the second model had a borderline significance of $p = 0.1$, for simplicity they were both omitted, which resulted in considering only the first model as the final model.

A prognostic index was defined as the linear predictor resulting from the final model. It was the result of the sum of covariate values for a particular patient, weighted by the corresponding estimated regression coefficients. According to this prognostic index, the population was split into the three categories, shown in Table 3.3.

Subsequently, the validation sample was also split into these three categories, and used to measure the predictive accuracy of the model by plotting the observed proportions of individuals with seizure recurrence, separated by treatment policy and stratified by risk recurrence levels. These plots were compared to the corresponding plots from the test sample.

The model and the Kaplan-Meier plots (Figures 3.1 and 3.2) give an in-

Prognostic index= 0 (Low risk of recurrence)	Prognostic index= 1 (Medium risk recurrence)	Prognostic index= 2 to 4 (High risk recurrence)
- A single seizure	- 2 to 3 seizures, normal EEG, no neurological abnormalities	- 1 seizure, abnormal EEG, neurological disorder
- A normal EEG	- 1 seizure, abnormal EEG, no neurological disorder	- 2-3 seizures, abnormal EEG, no neurological disorder
- Absence of neurological disorder	- 1 seizure, normal EEG, neurological disorder	- 2-3 seizures, normal EEG, neurological disorder - 4 or more seizures.

Table 3.3: Prognostic index as proposed by Kim et al. under a Cox's model. Table from Kim et al[27]

dication that, although the benefit of immediate treatment is not obvious for the low risk population, it does present a delay in the immediate seizure recurrence for the medium and high risk of recurrence populations. The results also support Marson et al.'s findings, regarding that there is no clinical benefit from immediate treatment in late risks, such as three to five years remission. The risk of seizure recurrence increases with the number of seizures at presentation, abnormal EEG, and the presence of neurological disorders.

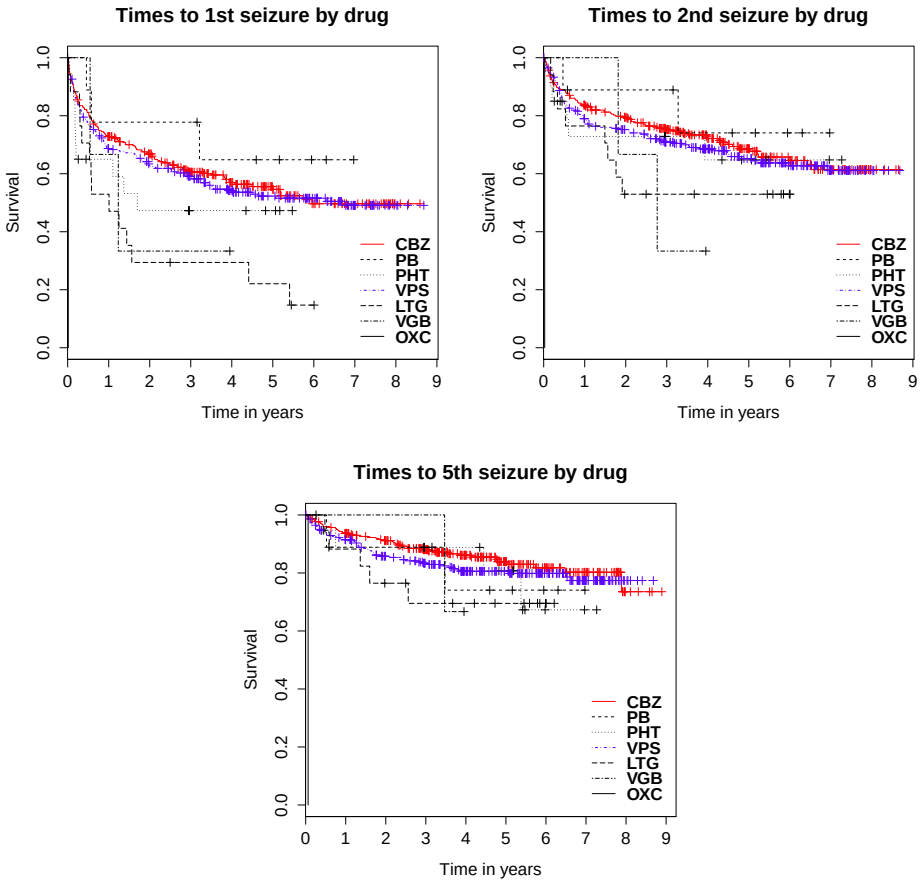


Figure 3.1: Kaplan-Meier survival plots for the first, second and fifth seizure times in years of occurrence, categorized by drug allocated to. Observe that the crosses denote a censored survival time.

In Figure 3.1, we can observe the Kaplan-Meier curves for the time from randomization to the times to first, second and fifth seizures, for each one of the different drugs. Observe that for the two drugs of main interest, CBZ and VPS, the differences between survival functions are not apparent until the fifth seizures. However, we must keep in mind that the number of patients who experience the fifth seizure is considerably less than the initial number of patients. In Figure 3.2 the difference between the survival curves is much more clear.

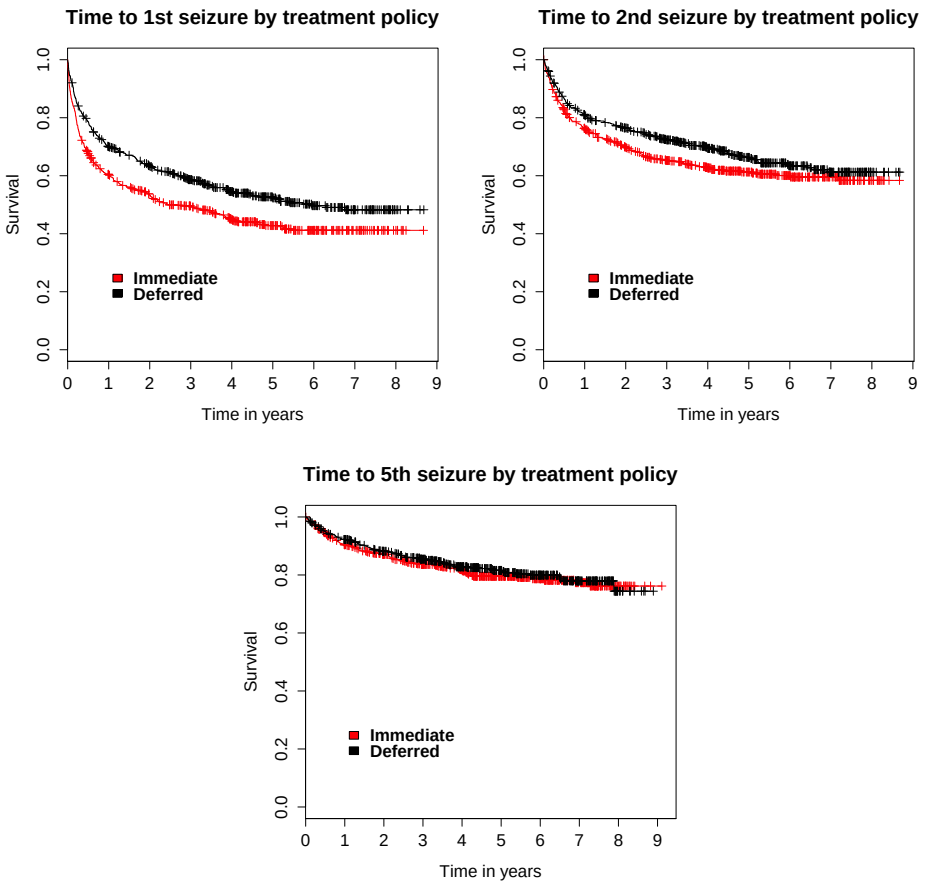


Figure 3.2: Kaplan-Meier survival plots for the first, second and fifth seizure times in years of occurrence, categorized by treatment allocation. Observe that the crosses denote a censored survival time.

3.3 Joint mixed models with and without cure fraction

The aim of this work is to include the information of each individual into the joint modelling of the number of seizure occurrences and the survival times. For the following models, some of the baseline covariates that were collected were age at randomization, sex, the patient's pre-randomization history, electoencephalogram (EEG). As part of the patient's history, their pre-randomization seizures are categorized as:

1. Tonic-Clonic seizures: which comprise tonic-clonic seizures only,

2. Secondary Tonic-Clonic seizures: these consist of partial seizures accompanied by second tonic-clonic seizures,

3. Generalized: which includes any types of generalized seizures, including any combinations of tonic-clonic and other generalized seizures,

4. Partial group: contains patients with simple or complex partial seizures only, and

5. Other: this group contains all the remaining seizures that could not be categorized in any of the other groups.

For each patient i $(i = 1, ..., n)$ in the study, there are recorded a pre-randomization event count X_i during a period of time u_i, a post-randomization time to first event Y_i which has either been observed or censored, and a set of covariates such as age, sex, type of drug administered and EEG recordings. From the nature of both the epilepsy disease and the study development, we assume that the underlying seizure rate λ_i for each individual is subjected to a possible modification due to the post-randomization effect ψ_i.

In the following sections we present two models proposed by Cowling (2003) and Rogers (2011).

3.3.1 Joint Mixed Model

The first joint model of interest to us was proposed and developed by Cowling ([11],[10]) for an epilepsy meta-analysis. It considers the point process to be an overdispersed homogeneous Poisson process with recurrence rate $\lambda_i \nu_i$ for every individual i, where ν_i is a random variable which partly explains the overdispersion observed in this kind of recurrent events. The pre-randomization covariates, Z_{1i},

are considered to have a log-link relationship with the underlying seizure rate λ_i, resulting in the model being a negative binomial Generalized Linear Model.

Given that the underlying count process is considered to be a Poisson process, the corresponding times to event are thought to follow an exponential distribution with rate $\lambda_i\psi_i\nu_i$. The treatment allocation is assumed to have a multiplicative effect on the underlying seizure rate, and it depends on a set of post-randomization covariates denoted by Z_{2i}. By considering a log-link of the covariates Z_{2i} on the multiplicative effect, this model can be then summarized in the following formulas.

$$\begin{aligned} f_X(x_i|\lambda_i, u_i, \nu_i) &= \frac{(\lambda_i u_i \nu_i)^{x_i} e^{-\lambda_i u_i \nu_i}}{x_i!}, \\ f_Y(y_i|\lambda_i, \psi_i, \nu_i) &= \lambda_i \psi_i \nu_i e^{\lambda_i \psi_i \nu_i y_i}, \\ g_\nu(\nu_i|\alpha) &= \frac{\alpha^\alpha \nu_i^{\alpha-1} e^{-\alpha\nu_i}}{\Gamma(\alpha)}, \end{aligned}$$

where

$$\begin{aligned} \lambda_i &= e^{\beta_1' z_{1i}}, \\ \psi_i &= e^{\beta_2' z_{2i}}. \end{aligned}$$

Observe that Z_{2i} is related to the post-randomization effect ψ_i by means of a log-link, and that it contains an intercept, a treatment indicator and a set of covariates and interactions considered relevant for the seizure rate change after randomization. Here, ν_i is assumed to follow a gamma distribution with mean 1 for identifiability purposes, and $\alpha > 0$ denotes the degree of heterogeneity in the population (the greater the value of α the smaller the heterogeneity will be).

For this model the derivation of the joint distributions for X_i and Y_i was found depending on whether Y_i was observed or censored. Let δ_i be the indicator which takes value one when the seizure time is observed, and zero when it is censored;

then for the case when the seizure time is observed, we have

$$\begin{aligned} f(x_i, y_i|u_i, \lambda_i, \psi_i, \alpha) &= \int_0^\infty f_X(x_i|u_i, \lambda_i, \nu_i) f_Y(y_i|\lambda_i, \psi_i, \nu_i) g_\nu(\nu_i|\alpha) d\nu_i \\ &= \frac{\alpha^\alpha \lambda_i^{x_i+1} u_i^{x_i} \psi_i}{x_i! \Gamma(\alpha)} \int_0^\infty \nu_i^{x_i+\alpha} \exp(-\nu_i(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)) d\nu_i, \end{aligned}$$

where the integrand in the last expression can be reformulated to take the form of a Gamma distribution, integrating to unity. Hence, the joint distribution is given by

$$\frac{\Gamma(x_i + \alpha + 1)\alpha^\alpha (\lambda_i u_i)^{x_i} \lambda_i \psi_i}{x_i! \Gamma(\alpha)(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^{x_i+\alpha+1}}. \tag{3.2}$$

For the case when Y_i is censored, the survivor function of the seizure time is $S_Y(y_i|\lambda_i, \psi_i, \nu_i) = \exp(-\lambda_i \psi_i \nu_i y_i)$. The joint distribution can now be found in a similar way to how equation 3.2 was computed. By using the survivor function for Y_i, the joint distribution is found to be of the form

$$\begin{aligned} P(X_i = x_i, Y_i > y_i; u_i, \lambda_i, \quad \psi_i \quad , \alpha) &= \int_0^\infty f_X(x_i|u_i, \lambda_i, \nu_i) S_Y(y_i|\lambda_i, \psi_i, \nu_i) g_\nu(\nu_i|\alpha) d\nu_i \\ &= \frac{\alpha^\alpha \lambda_i^{x_i} u_i^{x_i}}{x_i! \Gamma(\alpha)} \int_0^\infty \nu_i^{x_i+\alpha-1} \exp(-\nu_i(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)) d\nu_i \\ &= \frac{\Gamma(x_i + \alpha)\alpha^\alpha (\lambda_i u_i)^{x_i}}{x_i! \Gamma(\alpha)(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^{x_i+\alpha}}, \end{aligned} \tag{3.3}$$

where the expression in the integrand takes, once again, the form of a Gamma distribution and can be integrated out.

For this model, the marginal distributions for the seizure counts and the seizure recurrence time are found to be Negative Binomial and Lomax respectively. More specifically, the distribution for the X_i is a Negative Binomial with parameters α and $\alpha/(\lambda_i u_i + \alpha)$,

$$f(x_i|u_i, \lambda_i, \alpha) = \frac{\Gamma(x_i + \alpha)}{x_i! \Gamma(\alpha)} \left(\frac{\alpha}{\lambda_i u_i + \alpha}\right)^\alpha \left(\frac{\lambda_i u_i}{\lambda_i u_i + \alpha}\right)^{x_i}; \tag{3.4}$$

where the parameter ψ_i is not involved, since it is regarded as the treatment multiplicative change in the seizure rate, which is only available after randomization.

The corresponding marginal distribution for the post-randomization seizure recurrence time Y_i can be found to be of the form:

$$f(y_i|\lambda_i, \psi_i, \alpha) = \lambda_i \psi_i \left(\frac{\alpha}{\alpha + \lambda_i \psi_i y_i} \right)^{\alpha+1}. \tag{3.5}$$

3.3.2 Joint Mixed Model with a Cure Fraction

In Rogers and Hutton's work (Rogers 2011, Rogers and Hutton 2012) it is assumed that an individual i experiences seizures according to a Poisson process with rate $\lambda_i \nu_i$, where λ_i is a function of the baseline covariates, and ν_i is the frailty term. They denote by X_i the pre-randomization event count during a period u_i, T_{1i} and T_{2i} are the times for the first and second post-randomization seizure times, and set $Y_{1i} = T_{1i}$ and $Y_{2i} = T_{2i} - T_{1i}$, the gap times between post-randomization seizures. As a result of discussions with the clinicians, it was determined that the number of days at risk before randomization, u_i, would be adjusted so that, if t_i is the time of the first seizure (pre-randomization) and T_i is the time of randomization for individual i, it would take the following values

$$u_i = \max(182, |T_i - t_i|).$$

In this work, the model states that $X_i|\nu_i$ has a Poisson distribution with parameters $\lambda_i u_i \nu_i$, while the gap times follow an Exponential distribution with the same rate. Note that the unconditional probability of X_i, $f_X(x_i; \lambda_i, u_i, \alpha)$, is distributed as a Negative Binomial. The randomization effect was considered, as in Cowling et al. (2006), to be multiplicative and denoted by ν_i, which is distributed as $\nu_i \sim Gamma(\alpha, \alpha)$. Additionally, the post-randomization gap times Y_{1i} and Y_{2i} are both Exponentially distributed with rate $\lambda_i \psi_i \nu_i$, and the two survival times are

independent given ν_i. Finally we introduce the cure rates p_{1i} and p_{2i}, which depend on κ_1 and κ_2. Here κ_i represents the rate of increase or decrease of seizures relative to the susceptible proportion in the reference group, and p_{1i} is the probability that individual i has a first seizure, and p_{2i} is the probability that individual i has a second seizure. All this can be stated in the equations that follow:

$$\begin{aligned} f_{X|\nu}(x_i|\nu_i;\lambda_i,u_i) &= \frac{(\lambda_i u_i \nu_i)^{x_i} e^{\lambda_i u_i \nu_i}}{x_i!}, \\ f_{Y_1,Y_2|\nu}(y_{1i},y_{2i}|\nu_i;\lambda_i,\psi_i,p_{1i},p_{2i}) &= p_{1i}p_{2i}(\lambda_i\psi_i\nu_i)^2 e^{-\lambda_i\psi_i\nu_i(y_{1i}+y_{2i})}, \\ f_{\nu}(\nu_i;\alpha) &= \frac{\alpha^\alpha \nu_i^{\alpha-1} e^{-\alpha\nu_i}}{\Gamma(\alpha)}, \end{aligned}$$

where

$$\begin{aligned} \lambda_i &= \exp(\beta_1' z_{1i}), \quad \psi_i = \exp\left(\beta_2' z_{2i}\right), \\ p_{1i} &= \frac{\exp\left(\kappa_1'\omega_{1i}\right)}{1+\exp\left(\kappa_1'\omega_{1i}\right)}, \\ p_{2i} &= \frac{\exp\left(\kappa_2'\omega_{2i}\right)}{1+\exp\left(\kappa_2'\omega_{2i}\right)}, \end{aligned}$$

$(\beta_1, \beta_2, \kappa_1, \kappa_2)$ are regression paramenters, and z_{1i}, z_{2i}, ω_{1i}, ω_{2i} are vectors of covariates, not necessarily distinct. The unconditional joint distribution of the Y_{ij}, $j = 1, 2$, is a bivariate Lomax distribution, with each of the marginal Y_{ij} having a univariate Lomax distribution with shape and scale parameters α and $\alpha/\lambda_i\psi_i$ respectively. This results in an accelerated failure time model with a Lomax baseline distribution.

3.4 Model comparisons

In this section we will compare the predictions of the MESS epilepsy data set under the four principal models described in the two previous sections. For this section,

as well as for the residual analysis sections in Chapter 5, we will consider four models for the epilepsy data. The first two will be Cox proportional hazards models, one considering the set of covariates proposed by Kim et al.[27] and the second considering the set of covariates that will be used for the other two models, the joint model proposed by Cowling[10],[11] and the joint model with cure fraction proposed by Rogers[39],[41].

We will first replicate in the possible measure the results given by Kim et al., contrast its predictions to the ones provided by the joint and joint with cure fraction models for the same set of covariates, and at last, contrast the Cox models with the latter two under another set of covariates.

The approach used in the paper by Kim et al. was a Cox model, that considered the three covariates, EEG outcome, neurological disorder and the number of seizures pre-randomization, transformed by a logarithm function. From fitting such a model, we obtain that the regression coefficient estimates are as they appear in Table 3.4, where our estimated regression coefficients, their 95% confidence intervals and their respective p-values are contrasted with the corresponding Kim et al's findings. There is a relevant observation that must be made when considering the results that will be obtained here. While in the paper by Kim et al. the sample is split into two groups, 885 of patients were assigned as a test sample, 535 patients were used for validation, we do not split it so, as we do not carry out a cross-validation analysis. We consider the whole sample and complement the missing data with the mean values obtained from the remaining values of the population.

	$\hat{\beta}$	$\exp(\hat{\beta})$	$\exp(\hat{\beta}_{Kim})$	p-value	p-value$_{Kim}$
Neurological disorder	0.377	$1.46(1.35, 1.57)$	$1.35(1.07, 1.72)$	$< 2 \times 10^{-16}$	0.013
Abnormal EEG	0.438	$1.55(1.22, 1.97)$	$1.54(1.27, 1.86)$	3×10^{-4}	$< 1 \times 10^{-4}$
Log(No. of seizures)	0.287	$1.33(1.15, 1.55)$	$1.56(1.42, 1.72)$	2×10^{-4}	$< 1 \times 10^{-4}$

Table 3.4: Contrast between Kim et al.'s([27]) model estimates and our replication of the model.

Observe that the estimates are very similar, and the three covariates considered are deemed significant for the epilepsy recurrence time. The greater width of the confidence intervals for our reproduction of the fit could be due to the presence of a larger heterogeneity in the population when considering all patients, and not omitting individuals with incomplete data. We will now fit the joint and joint with cure fraction models to the same covariates, where the neurological disorder outcome is considered as a covariate for the underlying recurrence rate, λ_i, whilst the EEG outcome, the treatment allocation and first order interactions are considered for the post-randomization seizure rate modifier, ψ_i, and the cure fraction coefficient, κ. Observe that a fundamental difference between these models and the Cox model lies in the use of the number of pre-randomization seizures as a variable and not merely as a fixed covariate. We try to mirror Kim et al.'s considerations, and fit the models to the data stratified by treatment allocation, and do not consider interactions of the covariates.

		Cowling	SE	Rogers	SE
α		1.095	0.045	1.113	0.046
λ	Intercept	-4.671	0.032	-4.658	0.032
	NDL	0.261	0.105	0.247	0.105
ψ	Intercept	-2.293	0.077	-0.467	0.099
	Trt	-0.537	0.090	-0.652	0.119
	EEG	0.260	0.090	-0.328	0.122
	NDL	0.420	0.153	-0.337	0.209
κ	Intercept			0.012	0.087
	EEG			0.700	0.146
	NDL			0.894	0.329
	-Log-likelihood	8881.246		8634.132	

Table 3.5: Joint (Cowling) and joint with cure fraction (Rogers) models' estimates and standard errors, under the set of covariates proposed by Kim et al.[27].

From Table 3.5 it can be observed that, under the set of covariates proposed by Kim et al.[27] and considering the treatment allocation as an additional covariate, all the covariates of interest are found to be statistically significant for the pre and post-randomization rates, as well as for the cure fraction. From the values of α, the homogeneity in the population is estimated to be very similar between the Cowling and Rogers models, albeit the model with a cure fraction estimates a slightly more homogeneous population. Whilst the pre-randomization estimates do not present a noticeable difference between one another, the estimated EEG and the neurological disorder (NDL) coefficients produce an opposite effect on the seizure rate ψ between the two models. The existence of the cure fraction is supported by the difference between loglikelihood values, where we observe that the joint model with a cure fraction provides a better fit for the data, and that in the estimation of the cure fraction, both the neurological disorder and EEG outcomes are significant. The difficulty of considering the NDL covariate, however, lies in the fact that the covariate itself comprises three types of abnormality covariates (neurological deficit or impairment, delayed development and slow wave abnormality), and their joint outcomes are recorded for 132 of the 1425 initial number of patients, thus constituting 10.20% of the population. For this reason, we consider the alternative set of covariates used for the model fits shown in Table 3.6, shown below.

	Cox	SD	Cowling	SD	Rogers	SD
α			2.096	0.116	2.060	0.113
Intercept			-4.131	0.085	-4.129	0.085
Age			-0.005	0.002	-0.005	0.002
T-C			-1.085	0.094	-1.075	0.095
2nd.T-C			-0.697	0.097	-0.690	0.098
Intercept			-5.304	0.564	-4.046	0.615
Age	-0.001	0.002	0.004	0.003	-0.001	0.004
Trt	0.310	0.278	1.179	0.532	0.702	0.708
T-C	-0.247	0.268	2.945	0.569	3.514	0.626
2nd.T-C	-0.379	0.277	2.322	0.576	2.565	0.638
Trt*T-C	-0.375	0.282	-1.367	0.532	-1.010	0.715
Trt*2nd.T-C	-0.449	0.292	-1.446	0.542	-0.942	0.731
EEG	0.075	0.277	0.059	0.554	0.652	0.712
Trt*EEG	-0.548	0.161	-0.808	0.199	-0.860	0.281
EEG*T-C	0.356	0.287	0.410	0.556	-0.514	0.720
EEG*2nd.T-C	0.829	0.297	0.820	0.566	0.481	0.737
Log(No.seiz)	0.376	0.052				
κ					0.004	0.076
-Log-likelihood			7031.643		6861.389	

Table 3.6: Coefficient estimates for the Cox, Cowling and Rogers models.

In Table 3.6 we compare the Cox proportional hazards model, Cowling and Rogers models respectively, displaying their corresponding standard deviations and in the case of the joint models, the log-likelihood value. Since the Cox model only considers the number of seizures pre-randomization as a fixed covariate, the estimated coefficient for the logarithm of such number is included only for this model in the table. For this analysis, however, we do not stratify by treatment allocation but

rather consider an additional covariate for all three models. The reference population consists of patients at 30 years of age, under deferred treatment and presenting partial seizures only, with a normal EEG outcome.

Although the age of the patients at randomization appears to be non significant or borderline significant for both pre and post-randomization rates under all models considered, we keep this covariate for clinicians have deemed it relevant for the seizure recurrence prediction. The type of epilepsy covariates (T-C and 2nd.T-C) maintain very similar estimated coefficient values under Cowling and Rogers models for the seizure rate pre-randomization, and remain significant for both the pre and post-randomization seizure rates. It is noticeable that, in contrast, the T-C and 2nd. T-C seizure types are borderline significant for the Cox model, as well as almost all their first-order interactions with EEG outcome and treatment allocation. Only the interaction of EEG abnormality and 2nd. Tonic-Clonic seems relevant, for the Cox model as well as for the Cowling model, although this is not so for the joint model with the cure fraction. Indeed, for the Cox model all factors (T-C, 2nd.T-C, EEG and treatment) are borderline significant, it is the first-order interactions that have an effect on the seizure recurrence rate, along with the number of seizures pre-randomization.

The main difference between the models, in conclusion, lies in the fact that whilst the joint models show that the two types of epilepsy have a relevant contribution to the seizure rates, consistently for pre and post-randomization rates, the Cox model only consider their interactions with the treatment allocation and the EEG outcome to be of no significance. The EEG outcome factor contributes a significant change only under the joint model with a cure fraction, but the EEG interaction with the 2nd. tonic-clonic seizure provides a non-significant contribution only for this same model. The number of seizures pre-randomization is a very relevant covariate for the Cox model, whilst the estimated value of the cure fraction covariate κ appears to be negligible in relation with its standard deviation. This translates to

nearly half of the population not experience a seizure post-randomization. Indeed when $\kappa = 0$ the cure fraction of the population is 50%, and from the value of the standard deviation the estimated cure fraction could vary between 48.2% and 51.9%. The difference between log-likelihoods between the two models, however, shows an improvement of the model fit for the data under the assumption of the existence of a cure fraction of the population.

Chapter 4

Extensions of the Joint Model with and without Cure Fraction

4.1 Truncated Joint Model

In previous chapters we have addressed models proposed by Cowling (Cowling, 2003) and Rogers (Rogers, 2011), which consider an independent seizure process for each patient with and without a cure rate respectively. These models consider the seizure count to be Poisson distributed, taking into account patients who have any number of seizures, including those who have had no seizures. A patient is diagnosed with an epilepsy syndrome only when he has presented at least two unprovoked epileptic seizures, and in particular, the population under study have recorded at least one seizure. This leads us to propose a more realistic version of the previous models. Let us consider a joint probability model for the number of pre-randomization and post-randomization times of seizures, conditioned to have at least one seizure, i.e. $X_i > 0$ for every individual i. Then the zero-truncated model will be of the form

$$P(X_i = x_i, Y_i = y_i | X_i > 0) = \frac{P(X_i = x_i, Y_i = y_i)}{P(X_i > 0)} = \frac{P(X_i = x_i, Y_i = y_i)}{1 - P(X_i = 0)}.$$

In the following chapter we produce the derivation of the log-likelihood and its corresponding derivatives when the underlying point process is considered to be a zero-truncated Poisson-Gamma mixture model. Under this model, each patient is assumed to have at least one epileptic seizure pre-randomization, with seizure recurrence following a Poisson process with an individual seizure rate.

4.1.1 Derivation of the Truncated Joint Distribution

Let us consider the seizure count pre-randomization X_i to be zero truncated Poisson distributed with an underlying seizure rate $\lambda_i\nu_i$ over a period of time u_i, and considering a frailty term ν_i which follows a Gamma distribution with mean 1. The post-randomization effect ψ_i has a multiplicative effect on the seizure rate, and both λ_i and ψ_i depend on the covariates of individual i with Z_{1i} as the explanatory variables pre-randomization, and Z_{2i} as the treatment covariate for the multiplicative factor. The post-randomization times to first seizure Y_i are assumed to follow an exponential distribution with rate $\lambda_i\psi_i\nu_i$, and we will indicate whether or not it is a censored observation by means of the indicator δ_i. This means that δ_i will take a unit value if the time to first seizure is observed, and zero otherwise.

Zero Truncated Joint distribution with Y_i Observed

From the previous chapter in Section 3.3.1, it was shown in equation 3.2 that the joint distribution function of X_i and Y_i, when the seizure time is observed, is of the form

$$f_{X_i,Y_i}(x_i, y_i; u_i, \lambda_i, \psi_i, \alpha) = \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i}{x_i!\Gamma(\alpha)(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^{x_i+\alpha+1}}.$$

In order to find the corresponding zero-truncated joint distribution of X_i and Y_i, consider that the marginal distribution of X_i is a Negative Binomial of the form shown in equation 3.4. For simplicity, denote by ζ_i the expression $\zeta_i = \lambda_i u_i + \lambda_i\psi_i y_i + \alpha$, and by η_i the expression $\eta_i = \lambda_i u_i + \alpha$. This leads to the joint

distribution expression:

$$\begin{aligned} f_{X_i>0,Y_i}(x_i, y_i) &= \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i}{x_i!\Gamma(\alpha)\zeta_i^{x_i+\alpha+1}} \Big/ \left[1-\left(\frac{\alpha}{\lambda_i u_i+\alpha}\right)^\alpha\right] \\ &= \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{x_i+\alpha+1}[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]}. \quad (4.1) \end{aligned}$$

We are now interested in finding the corresponding marginal distributions of the variables of interest. The marginal distribution of X_i, in contrast to its non zero-truncated version, has an integral over Y_i, which will not separate in an expression multiplied by a gamma distribution that integrates to one. The marginal distribution for the pre-randomization seizure count can be found by the following process:

$$\begin{aligned} f_{X_i>0}(x_i;\alpha,\lambda_i) &= \int_0^\infty f_{X_i>0,Y_i}\left(x_i, y_i|\alpha,\lambda_i,\psi_i\right) dy_i \\ &= \int_0^\infty \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{x_i+\alpha+1}[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]} dy_i. \end{aligned}$$

Observe that only the expression $(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{x_i+\alpha+1}$ depends on y_i, hence we have

$$= \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]} \int_0^\infty (\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{-x_i-\alpha-1} dy_i,$$

from which the integral is found, and is evaluated at the limits thus obtaining

$$\begin{aligned} &= \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]} \\ &\quad \times \left(\frac{1}{-(x_i+\alpha)(\lambda_i\psi_i)}\right)\left[(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{-x_i-\alpha}\right]|_0^\infty \\ &= \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]} \\ &\quad \times \left(\frac{1}{-(x_i+\alpha)(\lambda_i\psi_i)}\right)\left[0-(\lambda_i u_i+\alpha)^{-x_i-\alpha}\right]. \end{aligned}$$

By re-arranging the terms, we obtain the distribution

$$= \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha](x_i+\alpha)(\lambda_i\psi_i)(\lambda_i u_i+\alpha)^{x_i+\alpha}},$$

which can be re-written in the simpler notation

$$= \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}}{x_i!\Gamma(\alpha)(x_i+\alpha)\eta_i^{x_i}(\eta_i^\alpha-\alpha^\alpha)}. \tag{4.2}$$

In the converse fashion, we seek to obtain the cumulative distribution function of X_i. Let us consider α as an integer number, then, by computing the sum of the terms given by Equation (4.2) we obtain

$$\begin{aligned} F_{X_i}(x_i;\alpha,\lambda\psi) &= \sum_{w=1}^{x_i} f_{x_i}(w;\alpha,\lambda_i,\psi_i) \\ &= \sum_{w=1}^{x_i} \frac{\Gamma(w+\alpha+1)\alpha^\alpha(\lambda_i u_i)^w}{w!\Gamma(\alpha)(w+\alpha)\eta_i^w(\eta_i^\alpha-\alpha^\alpha)}, \end{aligned}$$

and since, for α an integer number, we have $\Gamma(\alpha)=(\alpha-1)!$ and $\Gamma(w+\alpha+1)=(w+\alpha)!$, the expression takes the form

$$= \sum_{w=1}^{x_i} \frac{\alpha^\alpha}{\eta^\alpha-\alpha^\alpha}\binom{w+\alpha-1}{w}\left(\frac{\lambda_i u_i}{\eta_i}\right)^w.$$

Let us now factorize the terms that do not depend on the variable of interest, and complete the sum, to obtain the partial series of terms starting from $w=0$:

$$= \frac{\alpha^\alpha}{\eta_i^\alpha-\alpha^\alpha}\left[\sum_{w=0}^{x_i}\binom{w+\alpha-1}{w}\left(\frac{\lambda_i u_i}{\eta_i}\right)^w - 1\right],$$

and since $\eta_i=\lambda_i u_i+\alpha$, we complement the terms in the sum to obtain the form of

negative binomial distribution

$$= \frac{\alpha^\alpha}{\eta_i^\alpha - \alpha^\alpha}\left[\left(\frac{\alpha}{\eta_i}\right)^{-\alpha}\sum_{w=0}^{x_i}\binom{w+\alpha-1}{w}\left(\frac{\lambda_i u_i}{\eta_i}\right)^w\left(\frac{\alpha}{\eta_i}\right)^\alpha - 1\right]. \quad (4.3)$$

Observe that the partial series shown in Equation (4.3) constitutes the cumulative negative binomial distribution function evaluated at $X_i = x_i$. Consequently when x_i tends to infinity, this sum converges to unity and the cumulative distribution of X_i for the zero-truncated case is such that

$$\begin{aligned}
\lim_{x_i\to\infty} F_{X_i}(x_i;\alpha,\lambda\psi) &= \lim_{x_i\to\infty}\frac{\alpha^\alpha}{\eta_i^\alpha-\alpha^\alpha}\left[\left(\frac{\alpha}{\eta_i}\right)^{-\alpha}\sum_{w=0}^{x_i}\binom{w+\alpha-1}{w}\left(\frac{\lambda_i u_i}{\eta_i}\right)^w\left(\frac{\alpha}{\eta_i}\right)^\alpha - 1\right] \\
&= \frac{\alpha^\alpha}{\eta_i^\alpha-\alpha^\alpha}\left[\left(\frac{\eta_i}{\alpha}\right)^\alpha - 1\right] \\
&= \frac{\alpha^\alpha\eta_i^\alpha}{(\eta_i^\alpha-\alpha^\alpha)\alpha^\alpha} - \frac{\alpha^\alpha}{\eta_i^\alpha-\alpha^\alpha} \\
&= \frac{\eta_i^\alpha}{\eta_i^\alpha-\alpha^\alpha} - \frac{\alpha^\alpha}{\eta_i^\alpha-\alpha^\alpha} = 1. \quad (4.4)
\end{aligned}$$

Let us remember that $\eta_i = \lambda_i u_i + \alpha$, then $\eta_i^\alpha - \alpha^\alpha \geq 0$, and the density of X_i shown in Equation (4.2) is non-negative for all $X_i \in \mathbb{N}$. This result joint with the convergence to unity found in equation (4.4) shows that the zero-truncated function of X_i is indeed a density function over the space of the natural numbers.

Let us now consider the marginal distribution of Y_i for the zero-truncated model. For α belonging to the natural numbers, the marginal distribution of Y_i is found by summing the joint distribution given in (4.1), over all the possible values of X_i, resulting in the following expression.

$$\begin{aligned}
f_{Y_i}(y_i;\alpha,\lambda_i,\psi_i) &= \sum_{x_i=1}^{\infty} f_{X_i>0,Y_i}(x_i,y_i|\alpha,\lambda_i,\psi_i) \\
&= \sum_{x_i=1}^{\infty}\frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{x_i+\alpha+1}[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]}.
\end{aligned}$$

We proceed to multiply and divide by α in order to obtain the binomial combinatorial coefficient. In a second step, we complete the sum lower limit by adding and

substracting a term, producing a partial series which includes the value at $x_i = 0$

$$\begin{aligned} &= \sum_{x_i=1}^{\infty} \frac{\Gamma(x_i+\alpha+1)\alpha^{\alpha}(\lambda_i u_i)^{x_i}\lambda_i\psi_i\eta_i^{\alpha}}{x_i!\Gamma(\alpha)\zeta_i^{x_i+\alpha+1}(\eta_i^{\alpha}-\alpha^{\alpha})} \times \frac{\alpha}{\alpha} \\ &= \frac{\alpha^{\alpha+1}\lambda_i\psi_i\eta_i^{\alpha}}{(\eta_i^{\alpha}-\alpha^{\alpha})\zeta_i^{\alpha+1}} \left\{ \sum_{x_i=0}^{\infty} \binom{x_i+\alpha}{\alpha} \left(\frac{\lambda_i u_i}{\zeta_i}\right)^{x_i} - 1 \right\}. \end{aligned}$$

We now multiply and divide the sum in the equation by $(1-\lambda_i u_i/\zeta_i)^{\alpha}$, since $\zeta_i = \lambda_i u_i + \lambda_i\psi_i y_i + \alpha$, thus completing the form of a binomial distribution function

$$= \frac{\alpha^{\alpha+1}\lambda_i\psi_i\eta_i^{\alpha}}{(\eta_i^{\alpha}-\alpha^{\alpha})\zeta_i^{\alpha+1}} \left\{ \left(1-\frac{\lambda_i u_i}{\zeta_i}\right)^{-\alpha} \sum_{x_i=0}^{\infty} \binom{x_i+\alpha}{\alpha} \left(\frac{\lambda_i u_i}{\zeta_i}\right)^{x_i} \left(1-\frac{\lambda_i u_i}{\zeta_i}\right)^{\alpha} - 1 \right\},$$

and since $\left(\frac{\lambda_i u_i}{\zeta_i}\right) < 1$ and $\left(1-\frac{\lambda_i u_i}{\zeta_i}\right) < 1$, the sum converges to 1 and we obtain the expression

$$\begin{aligned} &= \frac{\alpha^{\alpha+1}\lambda_i\psi_i\eta_i^{\alpha}}{(\eta_i^{\alpha}-\alpha^{\alpha})\zeta_i^{\alpha+1}} \left[\left(1-\frac{\lambda_i u_i}{\zeta_i}\right)^{-\alpha} - 1 \right] \\ &= \frac{\alpha^{\alpha+1}\lambda_i\psi_i\eta_i^{\alpha}}{(\eta_i^{\alpha}-\alpha^{\alpha})\zeta_i^{\alpha+1}} \left[\left(\frac{\zeta_i}{\lambda_i\psi_i y_i+\alpha}\right)^{\alpha} - 1 \right]. \end{aligned} \tag{4.5}$$

Observe that, since $\lambda_i > 0$ and $u_i > 0$ for all $i \in \{1, 2, 3, ..., 1334\}$, we have that

$$\frac{\zeta_i}{\lambda_i\psi_i y_i+\alpha} = \frac{\lambda_i u_i + \lambda_i\psi_i y_i + \alpha}{\lambda_i\psi_i y_i+\alpha} > 1.$$

From this we conclude that the marginal density function of Y_i obtained in equation (4.5) is non-negative for all $y_i \geq 0$.

The corresponding cumulative distribution function for Y_i can be found by integrating the density function obtained in equation (4.5). Let $\gamma \geq 0$, hence distri-

bution function of Y_i evaluated at γ is given by

$$\begin{aligned} F_{Y_i}(\gamma;\alpha,\lambda_i,\psi_i) &= \int_0^{\gamma} f_{y_i}(y;\alpha,\lambda_i,\psi_i)dy \\ &= \int_0^{\gamma} \frac{\alpha^{\alpha+1}\lambda_i\psi_i\eta_i^{\alpha}}{(\eta_i^{\alpha}-\alpha^{\alpha})\zeta_i^{\alpha+1}}\left[\left(\frac{\zeta_i}{\lambda_i\psi_i u_i+\alpha}\right)^{\alpha}-1\right]dy, \end{aligned}$$

where the expression in the integral depends on y through the term ζ_i, and by splitting the integral by the sum of the two terms, we obtain

$$\begin{aligned} &= \int_0^{\gamma} \frac{\zeta_i^{\alpha}\lambda_i\psi_i\eta_i^{\alpha}\alpha}{\zeta_i^{\alpha+1}(\eta_i^{\alpha}-\alpha^{\alpha})}\left(\frac{\alpha}{\lambda_i\psi_i y+\alpha}\right)^{\alpha}dy - \int_0^{\gamma}\frac{\alpha^{\alpha+1}\lambda_i\psi_i\eta_i^{\alpha}}{\zeta_i^{\alpha+1}(\eta_i^{\alpha}-\alpha^{\alpha})}dy \\ &= \frac{\alpha\eta_i^{\alpha}}{\eta_i^{\alpha}-\alpha^{\alpha}}\int_0^{\gamma}\frac{\lambda_i\psi_i}{\zeta_i}\left(\frac{\alpha}{\lambda_i\psi_i y+\alpha}\right)^{\alpha}dy - \frac{\alpha^{\alpha+1}\lambda_i\psi_i\eta_i^{\alpha}}{\eta_i^{\alpha}-\alpha^{\alpha}}\int_0^{\gamma}\zeta_i^{-\alpha-1}dy. \quad (4.6) \end{aligned}$$

Let us now remember, from Equation (3.5) from the previous chapter, that the Lomax distribution function has the form

$$f(y_i|\lambda_i,\psi_i,\alpha) = \lambda_i\psi_i\left(\frac{\alpha}{\alpha+\lambda_i\psi_i y_i}\right)^{\alpha+1},$$

hence we complete the term in the first integral of the equation (4.6) to obtain such a shape. The second integral in the expression can also be directly integrated and evaluated at the integration limits,

$$= \frac{\alpha\eta_i^{\alpha}}{\eta_i^{\alpha}-\alpha^{\alpha}}\int_0^{\gamma}\lambda_i\psi_i\left(\frac{\alpha}{\lambda_i\psi_i y+\alpha}\right)^{\alpha+1}\frac{\lambda_i\psi_i y+\alpha}{\alpha\zeta_i}du - \frac{\alpha^{\alpha+1}\lambda_i\eta_i^{\alpha}}{\eta_i^{\alpha}-\alpha^{\alpha}}\left(\frac{\zeta_i^{-\alpha}}{(-\alpha)\lambda_i\psi_i}\Big|_0^{y}\right).$$

Observe that, $\lambda_i\psi_i y+\alpha = \zeta_i - \lambda_i u_i$, hence the expression in the first integral can

be rewritten as

$$
\begin{aligned}
&= \frac{\eta_i^\alpha}{\eta_i^\alpha - \alpha^\alpha} \int_0^\gamma \lambda_i \psi_i \left(\frac{\alpha}{\lambda_i \psi_i y + \alpha} \right)^{\alpha+1} \left(\frac{\zeta_i - \lambda_i u_i}{\zeta_i} \right) du + \frac{\alpha^\alpha \eta_i^\alpha}{\eta_i^\alpha - \alpha^\alpha} \zeta_i^{-\alpha} \Big|_0^\gamma \\
&= \frac{\eta_i^\alpha}{\eta_i^\alpha - \alpha^\alpha} \left[\int_0^\gamma \lambda_i \psi_i \left(\frac{\alpha}{\lambda_i \psi_i y + \alpha} \right)^{\alpha+1} dy \right. \\
&\quad \left. - \int_0^\gamma \lambda_i \psi_i \left(\frac{\alpha}{\lambda_i \psi_i y + \alpha} \right)^{\alpha+1} \left(\frac{\lambda_i u_i}{\zeta_i} \right) du + \alpha^\alpha \zeta_i^{-\alpha} \Big|_0^y \right].
\end{aligned}
$$

The first integral of the equation now constitutes a Lomax distribution function, and as such, we find that the expression becomes

$$
= \frac{\eta_i^\alpha}{\eta_i^\alpha - \alpha^\alpha} \left\{ \left[1 - \left(\frac{\alpha}{\lambda_i \psi_i y + \alpha} \right)^\alpha \right] \Big|_0^y + \alpha^\alpha \zeta_i^{-\alpha} |_0^y - \lambda_i u_i \int_0^\gamma \frac{\lambda_i \psi_i}{\zeta_i} \left(\frac{\alpha}{\lambda_i \psi_i y + \alpha} \right)^{\alpha+1} dy \right\},
$$

and evaluating on the integration limits, we obtain

$$
\begin{aligned}
&= \frac{\eta_i^\alpha}{\eta_i^\alpha - \alpha^\alpha} \left\{ 1 - \left(1 + \frac{\lambda_i \psi_i}{\alpha} \gamma \right)^{-\alpha} - 1 + \left(1 + \frac{\lambda_i \psi_i}{\alpha} 0 \right)^{-\alpha} \right. \\
&\quad \left. + \alpha^\alpha \left[\frac{1}{(\lambda_i u_i + \lambda_i \psi_i \gamma + \alpha)^\alpha} - \frac{1}{(\lambda_i u_i + \alpha)^\alpha} \right] - \lambda_i u_i \int_0^\gamma \frac{\lambda_i \psi_i}{\zeta_i} \left(\frac{\alpha}{\lambda_i \psi_i y + \alpha} \right)^{\alpha+1} dy \right\}.
\end{aligned}
$$

From these equations we arrive at the final form of the zero-truncated distribution function

$$
\begin{aligned}
F_{Y_i}(y_i; \alpha, \lambda_i, \psi_i) &= \frac{\eta_i^\alpha}{\eta_i^\alpha - \alpha^\alpha} \left\{ 1 - \left(1 + \frac{\lambda_i \psi_i}{\alpha} \gamma \right)^{-\alpha} + \alpha^\alpha \left(\zeta_i^{-\alpha} - \eta_i^{-\alpha} \right) \right. \\
&\quad \left. - \lambda_i u_i \int_0^{y_i} \frac{\lambda_i \psi_i}{\zeta_i} \left(\frac{\alpha}{\lambda_i \psi_i u + \alpha} \right)^{\alpha+1} du \right\}. \qquad (4.7)
\end{aligned}
$$

We leave the zero-truncated cumulative distribution function expressed in terms of the integral as shown in Equation (4.7), where the integrand consists of a Lomax density function divided by a linear function of y. Such integral is not straightforward to compute, since integration by parts diminishes the power of one term but increases the power of the other term. One possible interpretation of such an integral, as γ tends to infinity, would constitute the expectation with respect to a

Lomax distribution of the form

$$\lim_{\gamma\to\infty}\int_0^{\gamma}\frac{\lambda_i\psi_i}{\zeta_i}\left(\frac{\alpha}{\lambda_i\psi_i y+\alpha}\right)^{\alpha+1}dy=\mathbb{E}\left[\frac{1}{\lambda_i u_i+\lambda_i\psi_i Y+\alpha}\right].$$

Although the closed form of the cumulative distribution of Y_i for a finite value y_i is desirable in order to obtain the survival function, this is left as a project for future work due to constraints on time. This does not restrain us, however, from computing the likelihood function of the joint distribution of X_i and Y_i, and from finding the appropriate covariate estimates under the model.

For the asymptotic behaviour of Equation (4.7), let us remember that Fubini-Tonelli's theorem states that, for two random variables Z and W with σ-finite spaces, and $g : Z \times W \to [0,\infty]$, without loss of generality if

$$\sum_Z\left(\int_W|g(z,w)|dw\right)<\infty,$$

then

$$\sum_Z\left(\int_W g(z,w)dw\right)=\int_W\left(\sum_Z g(z,w)\right)dw.$$

In Equation (4.4) we proved that the marginal distribution of X_i integrates to one, hence we have in particular that, for $f_{X_i>0,Y_i}(x_i,y_i)$ representing the zero-truncated joint density function of X_i and Y_i as defined in Equation(4.1), the following condition holds

$$\sum_{x_i=1}^{\infty}\left(\int_0^{\infty}|f_{X_i>0,Y_i}(x_i,y_i;\alpha,\lambda_i,\psi_i)|dy_i\right)=\sum_{x_i=1}^{\infty}\left(\int_0^{\infty}f_{X_i>0,Y_i}(x_i,y_i;\alpha,\lambda_i,\psi_i)dy_i\right)=1.$$

The first equality between the equations is derived from the fact that for all $x_i \in \mathbb{N}$ and $y_i \in \mathbb{R}^+$, the joint density is non-negative. Fubini-Tonelli's theorem then concludes that necessarily

$$\int_0^{\infty}\left(\sum_{x_i=1}^{\infty}f_{X_i>0,Y_i}(x_i,y_i;\alpha,\lambda_i,\psi_i)\right)dy_i=1,$$

and since $f_{Y_i}(y_i; \alpha, \lambda_i, \psi_i)$ is non-negative for all $y_i \in \mathbb{R}^+$, it is a density function of Y_i for each patient i in the population.

Zero Truncated Joint distribution with Y_i Censored

From equation 3.3 introduced in Section 3.3.1, the joint distribution for the case of censored seizure times (i.e. $\delta_i = 0$) was shown to be

$$f_{X_i,Y_i}(x_i, Y_i > y_i; u_i, \lambda_i, \psi_i, \alpha) = \frac{\Gamma(x_i + \alpha)\alpha^\alpha(\lambda_i u_i)^{x_i}}{x_i!\Gamma(\alpha)(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^{x_i+\alpha}}.$$

When dividing by the probability of the seizure count being strictly positive, we obtain the truncated distribution function

$$\begin{aligned} f_{X_i>0,Y_i}(x_i, Y_i > y_i; u_i, \lambda_i, \psi_i, \alpha) &= \frac{\Gamma(x_i + \alpha)\alpha^\alpha(\lambda_i u_i)^{x_i}}{x_i!\Gamma(\alpha)(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^{x_i+\alpha}} \Big/ \left[1 - \left(\frac{\alpha}{\lambda_i u_i + \alpha}\right)^\alpha\right] \\ &= \frac{\Gamma(x_i + \alpha)\alpha^\alpha(\lambda_i u_i)^{x_i}(\lambda_i u_i + \alpha)^\alpha}{x_i!\Gamma(\alpha)(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^{x_i+\alpha}[(\lambda_i u_i + \alpha)^\alpha - \alpha^\alpha]}. \end{aligned}$$

4.1.2 The Full Log-Likelihood and Derivatives

The likelihood function is the product of the density functions over all patients, using δ_i to denote whether their survival time has been observed or censored. Let us remember that the parameters of interest λ_i and ψ_i depend, in turn, on the covariate vectors β_1 and β_2 respectively in the form $\lambda_i = \exp(z_{1i}\beta_1)$ and $\psi_i = \exp(z_{2i}\beta_2)$. If x and y are respectively the sets of seizure counts and first seizure times post-

randomization for all n patients, we obtain the equation:

$$\begin{aligned}
\mathcal{L}(\alpha,\beta_1,\beta_2;x,y) &= \prod_{i=1}^{n}\left[\frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{x_i+\alpha+1}[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]}\right]^{\delta_i} \\
&\quad\times\left[\frac{\Gamma(x_i+\alpha)\alpha^\alpha(\lambda_i u_i)^{x_i}(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{x_i+\alpha}[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]}\right]^{1-\delta_i} \\
&= \prod_{i=1}^{n}\frac{\Gamma(x_i+\alpha)\alpha^\alpha(\lambda_i u_i)^{x_i}(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)(\lambda_i u_i+\lambda_i\psi_i y_i+\alpha)^{x_i+\alpha}[(\lambda_i u_i+\alpha)^\alpha-\alpha^\alpha]} \\
&\quad\times\left[\frac{(x_i+\alpha)\lambda_i\psi_i}{\lambda_i u_i+\lambda_i\psi_i y_i+\alpha}\right]^{\delta_i} \\
&\quad\text{and with the simplified notation, the equation is expressed as} \\
&= \prod_{i=1}^{n}\frac{\Gamma(x_i+\alpha)\alpha^\alpha(\lambda_i u_i)^{x_i}\eta_i^\alpha}{x_i!\Gamma(\alpha)\zeta_i^{x_i+\alpha}[\eta_i^\alpha-\alpha^\alpha]}\left[\frac{(x_i+\alpha)\lambda_i\psi_i}{\zeta_i}\right]^{\delta_i}.
\end{aligned}$$

By applying the logarithm to the likelihood function we obtain the log-likelihood function.

$$\begin{aligned}
\ell(\alpha,\beta_1,\beta_2;x,y) &= \log[\mathcal{L}(\alpha,\beta_1,\beta_2;x,y)] \\
&= \sum_{i=1}^{n}\{\log\Gamma(x_i+\alpha)-\log\Gamma(\alpha)-\log(x_i!)+\alpha\log(\alpha)+x_i\log(\lambda_i u_i) \\
&\quad+\alpha\log(\eta_i)-(x_i+\alpha)\log(\zeta_i)-\log(\eta_i^\alpha-\alpha^\alpha) \\
&\quad+\delta_i\left[\log(x_i+\alpha)+\log(\lambda_i)+\log(\psi_i)-\log(\zeta_i)\right]\}.
\end{aligned}$$

For the purpose of simplifying the expressions corresponding to the first and second derivatives of the log-likelihood with respect to α, β_1 and β_2, we show a middle step in the computations of the derivatives, with the hope of making the computations easier to follow to the reader. We take into consideration that, for any constant c,

$$\begin{aligned}
\frac{d\alpha^{\alpha+1}}{d\alpha} &= (1+\alpha)\alpha^\alpha+\alpha^{\alpha+1}\log(\alpha), \\
\frac{d(c+\alpha)^{\alpha+1}}{d\alpha} &= (1+\alpha)(c+\alpha)^\alpha+(c+\alpha)^{\alpha+1}\log(c+\alpha),
\end{aligned}$$

and in particular, for our definition of ζ_i, its derivation with respect to α yields:

$$\frac{d\zeta_i^{\alpha+1}}{d\alpha} = (1+\alpha)\zeta_i^{\alpha} + \zeta_i^{(}\alpha+1)\log(\zeta_i) = \zeta_i^{\alpha}[1+\alpha+\zeta_i\log(\zeta_i)].$$

For the derivations of the various expressions under β_1 and β_2, observe that from the definition of λ and ψ, their respective relevant derivatives are of the form:

$$\begin{aligned}
\frac{d\lambda}{d\beta_1} &= \frac{d\exp(\beta_1 z_1)}{d\beta_1} = z_1\exp(\beta_1 z_1) = z_1\lambda, \\
\frac{d\psi}{d\beta_2} &= \frac{d\exp(\beta_2 z_2)}{d\beta_2} = z_2\exp(\beta_2 z_2) = z_2\psi,
\end{aligned}$$

which lead to the respective derivatives of ζ_i with respect to β_1 and β_2 shown below,

$$\begin{aligned}
\frac{d\zeta_i}{d\beta_1} &= z_1\lambda_i(u_i+\psi_i y_i) = z_1\lambda_i\eta_i, \\
\frac{d\zeta_i}{d\beta_2} &= z_2\psi_i\lambda_i y_i.
\end{aligned}$$

From these considerations and notations, we are able to derive the corresponding first derivatives of the log-likelihood with respect to the parameters of interest.

$$\begin{aligned}
\frac{\partial\ell(\alpha,\beta_1,\beta_2;x,y)}{\partial\alpha} &= \sum_{i=1}^{n}\Bigg\{\frac{\partial\log\Gamma(x_i+\alpha)-\partial\log\Gamma(\alpha)}{\partial\alpha} + \log(\alpha) + \frac{\alpha}{\alpha} + \log(\eta_i) + \frac{\alpha}{\eta_i} \\
&\quad - \log(\zeta_i) - \frac{x_i+\alpha}{\zeta_i} - \frac{\alpha\eta_i^{\alpha-1}+\eta_i^{\alpha}\log(\eta_i)-\alpha^{\alpha}[1+\log(\alpha)]}{\eta_i^{\alpha}-\alpha^{\alpha}} \\
&\quad + \delta_i\left[\frac{1}{x_i+\alpha}-\frac{1}{\zeta_i}\right]\Bigg\} \\
&= \sum_{i=1}^{n}\Bigg\{\frac{\partial\log\Gamma(x_i+\alpha)-\partial\log\Gamma(\alpha)}{\partial\alpha} + \log(\alpha) + 1 + \log(\eta_i) - \log(\zeta_i) \\
&\quad + \frac{\alpha}{\eta_i} - \frac{x_i+\alpha-\delta_i}{\zeta_i} - \frac{\eta_i^{\alpha-1}[\alpha+\eta_i\log(\eta_i)]-\alpha^{\alpha}[1+\log(\alpha)]}{\eta_i^{\alpha}-\alpha^{\alpha}} \\
&\quad + \frac{\delta_i}{x_i+\alpha}\Bigg\}.
\end{aligned}$$

Observe that although the terms in the equation could be further grouped by the value of δ_i, it does not provide a greater advantage computationally for the R package *optim*. In the following derivatives, we therefore group the terms in the

equations in a similar way.

$$\begin{aligned}
\frac{\partial \ell(\alpha,\beta_1,\beta_2;x,y)}{\partial \beta_1} &= \sum_{i=1}^{n}\left\{\frac{x_i u_i z_1\lambda_i}{\lambda_i u_i}+\frac{\alpha z_1\lambda_i u_i}{\eta_i}-(x_i+\alpha)\frac{z_1\lambda_i+z_1\lambda_i\psi_i y_i}{\zeta_i}\right.\\
&\quad \left.-\frac{\alpha\eta_i^{\alpha-1}z_1\lambda_i u_i}{\eta_i^{\alpha}-\alpha^{\alpha}}+\delta_i\left[\frac{z_1\lambda_i}{\lambda_i}-\frac{z_1\lambda_i u_i+z_1\lambda_i\psi_i y_i}{\zeta_i}\right]\right\}\\
&\quad \text{by factorizing } z_1\lambda_i \text{ out of the equation, we obtain the expression}\\
&= \sum_{i=1}^{n} z_1\lambda_i\left\{\frac{x_i}{\lambda_i}\frac{\alpha u_i}{\eta_i}-\frac{(x_i+\alpha)(u_i+\psi_i y_i)}{\zeta_i}-\frac{\alpha\eta_i^{\alpha-1}u_i}{\eta_i^{\alpha}-\alpha^{\alpha}}\right.\\
&\quad \left.+\delta_i\left[\frac{1}{\lambda_i}-\frac{u_i+\psi_i y_i}{\zeta_i}\right]\right\}\\
&= \sum_{i=1}^{n} z_1\lambda_i\left\{\frac{x_i+\delta_i}{\lambda_i}+\frac{\alpha u_i}{\eta_i}-\frac{(x_i+\alpha+\delta_i)(u_i+\psi_i y_i)}{\zeta_i}-\frac{\alpha\eta_i^{\alpha-1}u_i}{\eta_i^{\alpha}-\alpha^{\alpha}}\right\}\\
\frac{\partial \ell(\alpha,\beta_1,\beta_2;x,y)}{\partial \beta_2} &= \sum_{i=1}^{n}\left\{-(x_i+\alpha)\frac{z_2\psi_i\lambda_i y_i}{\zeta_i}+\delta_i\left[\frac{z_2\psi_i}{\psi_i}-\frac{z_2\psi_i\lambda_i y_i}{\zeta_i}\right]\right\}\\
&\quad \text{by factorizing } z_2\psi_i, \text{ the expression becomes}\\
&= \sum_{i=1}^{n} z_2\psi_i\left\{-\frac{(x_i+\alpha)\lambda_i y_i+\delta_i\lambda_i y_i}{\zeta_i}+\frac{\delta_i}{\psi_i}\right\}\\
&= \sum_{i=1}^{n} z_2\psi_i\left\{-\frac{(\lambda_i y_i)(x_i+\alpha+\delta_i)}{\zeta_i}+\frac{\delta_i}{\psi_i}\right\}.
\end{aligned}$$

Observe that the partial derivatives of the log-likelihood with respect to β_1 and β_2 share a similar structure when the denominator is ζ_i. The main difference resides in the rest of the terms, which depend solely on λ_i or ψ_i respectively.

From these equations we obtain the second order partial derivatives. Although the computations are fairly straightforward, we omit various stages of algebraic manipulations, showing in the equations below, the simplified expressions of

the derivatives.

$$
\begin{aligned}
\frac{\partial^2 \ell(\alpha,\beta_1,\beta_2;x,y)}{\partial \beta_1^2} &= \sum_{i=1}^{n} z_1^T z_1 \lambda_i \left\{ \frac{x_i+\delta_i}{\lambda_i} + \frac{\alpha u_i}{\eta_i} - \frac{(u_i+\psi_i y_i)(x_i+\alpha+\delta_i)}{\zeta_i} - \frac{\alpha \eta_i^{\alpha-1} u_i}{\eta_i^\alpha - \alpha^\alpha} \right\} \\
&\quad + \sum_{i=1}^{n} z_1^T z_1 \lambda_i^2 \left\{ -\frac{x_i+\delta_i}{\lambda_i^2} - \frac{\alpha u_i}{\eta_i} + \frac{(u_i+\psi_i y_i)^2 (x_i+\alpha+\delta_i)}{\zeta_i^2} \right. \\
&\quad \left. - \frac{\alpha(\alpha-1)\eta_i^{\alpha-2} u_i^2 (\eta_i^\alpha - \alpha^\alpha) - (\alpha \eta_i^{\alpha-1} u_i)^2}{(\eta_i^\alpha - \alpha^\alpha)^2} \right\}, \\
\frac{\partial^2 \ell(\alpha,\beta_1,\beta_2;x,y)}{\partial \beta_2^2} &= \sum_{i=1}^{n} \left\{ z_2^T z_2 \psi_i \left[\frac{\delta_i}{\psi_i} - \frac{\lambda_i y_i (x_i+\alpha+\delta_i)}{\zeta_i} \right] \right. \\
&\quad \left. + z_2^T z_2 \psi_i^2 \left[\frac{(\lambda_i y_i)^2 (x_i+\alpha+\delta_i)}{\zeta_i^2} - \frac{\delta_i}{\psi_i^2} \right] \right\}, \\
\frac{\partial^2 \ell(\alpha,\beta_1,\beta_2;x,y)}{\partial \beta_2^2} &= \sum_{i=1}^{n} \left\{ \frac{\partial^2 \log\Gamma(x_i+\alpha) - \partial^2 \log\Gamma(\alpha)}{\partial \alpha^2} + \frac{1}{\alpha} + \frac{1}{\eta_i} - \frac{1}{\zeta_i} + \frac{\eta_i - \alpha}{\eta_i^2} \right. \\
&\quad - \frac{\zeta_i - x_i + \alpha - \delta_i}{\zeta_i^2} - \frac{\alpha^\alpha (1+\log(\alpha))^2 + \alpha^{\alpha-1}}{\eta_i^\alpha - \alpha^\alpha} \\
&\quad - \frac{(\alpha + \eta_i \log(\eta_i)) \eta_i^{\alpha-2} (\alpha - 1 + \eta_i \log(\eta_i)) + \eta_i^{\alpha-1} (\log(\eta_i) + 2)}{\eta_i^\alpha - \alpha^\alpha} \\
&\quad \left. + \left[\frac{\eta_i^{\alpha-1} (\alpha + \eta_i \log(\eta_i)) - \alpha^\alpha (1+\log(\alpha))}{\eta_i^\alpha - \alpha^\alpha} \right]^2 \right\}.
\end{aligned}
$$

These second order derivatives are used in the optimization process carried out in R. They provide a faster approximation to the maximum likelihood estimates, and can be directly used to obtain the standard errors for the corresponding covariate coefficient estimates. It is worth mentioning that the partial derivative functions of the log-likelihood were checked with numerical estimates found with R.

4.1.3 Maximum Likelihood Estimation

The fitting of this model to a given data set can be carried out by means of an optimization algorithm, such as the Newton-Raphson approximation method. For this process the log-likelihood may be found under an iterative process using a suitable set of initial values of α, β_1 and β_2, and the first (and possibly the second) derivatives of such log-likelihood function. For the joint and joint with cure fraction models, the initial values for the optimization method are found by means of fitting

a negative binomial generalized linear model to the count data alone. The regression coefficient estimates are used as the initial values for β_1, whilst a survival regression assuming a log-logistic distribution is used to find the initial values of β_2. For the truncated joint model, the initial values of the parameters have been fitted using the negative binomial regression estimates, and in a second stance, as the maximum likelihood estimates obtained from the joint model.

For the approximation of the term $\log\Gamma(x_i+\alpha)-\log\Gamma(\alpha)-\log(x_i!)$ present in the log-likelihood, we have compared the use of the Stirling formula, the *lgamma* function in R and the original set up shown in Rogers (Rogers, 2011), in which the term is broken into the term $\sum_{k=0}^{x_i-1}\log(\alpha+k)$. It was found that for some cases, the values of the sums in the method used by Rogers could cause the optimization to diverge. The Stirling formula alleviates this problem to a certain degree, but the optimization considering the function *lgamma* converges both in the epilepsy data set and in the simulations described in the goodness of fit chapter with a greater speed.

For the derivation of the corresponding expression,

$$\frac{\partial\log\Gamma(x_i+\alpha)}{\partial\alpha}-\frac{\partial\log\Gamma(\alpha)}{\partial\alpha},$$

we use the R function *digamma*, which computes the needed first-order derivative.

4.1.4 Application of the Truncated Joint Model to the MESS Data

In this section we fit the joint model, the joint model with a cure fraction and the joint zero-truncated model to the MESS epilepsy data set. The reference group is defined as individuals presenting partial seizures pre-randomization, a normal EEG and randomized to deferred treatment. A positive (negative) regression coefficient would indicate an increase (decrease) in seizure rate relative to the seizure rate of the reference group. Similarly, a positive estimate of κ would indicate an increase

in the susceptible proportion of patients, relative to the reference group. Finally, a greater value of the estimated α will indicate a large similarity between individuals, whilst a small value is an indicator of a large heterogeneity in the population.

Considering only the first time to seizure occurrence after randomization, we obtain the following Table 4.1, which shows the maximum likelihood estimates for the covariates considered in three models: joint model (Cowling), zero truncated joint model (Truncated Cowling) and the joint model with a cure fraction (Rogers). In this case we took the initial values for the optimization of the truncated model to be based on a combination of the estimate of α from the joint model, and the rest of the estimates are originated from a negative binomial regression.

		Cowling	SE	Truncated Cowling	SE	Rogers	SE
α		2.095	0.116	10.083	0.125	2.060	0.113
λ	Intercept	-4.131	0.085	-4.411	0.062	-4.129	0.085
	Age	-0.005	0.002	-0.012	0.001	-0.005	0.002
	T-C	-1.085	0.094	-1.335	0.076	-1.075	0.095
	2nd.T-C	-0.697	0.097	-0.738	0.073	-0.690	0.098
ψ	Intercept	-5.303	0.564	-7.219	1.214	-4.045	0.615
	Age	0.004	0.003	0.012	0.004	0.0001	0.004
	Trt	1.178	0.532	1.688	1.038	0.702	0.708
	T-C	2.945	0.569	3.896	1.215	3.513	0.626
	2nd.T-C	2.322	0.576	3.032	1.225	2.565	0.638
	Trt*T-C	-1.366	0.532	-1.367	1.031	-1.009	0.715
	Trt*2nd.T-C	-1.446	0.542	-1.655	1.036	-0.942	0.731
	EEG	0.059	0.554	0.927	1.093	0.652	0.712
	Trt*EEG	-0.808	0.199	-1.246	0.270	-0.860	0.281
	EEG*T-C	0.409	0.556	0.380	1.089	-0.514	0.720
	EEG*2nd.T-C	0.820	0.566	1.292	1.100	0.480	0.737
κ						0.004	0.076
	-Loglikelihood	7031.643		6799.428		6861.389	

Table 4.1: Cowling, Rogers and Truncated joint model maximum likelihood parameter estimates.

The frailty term parameter α takes a larger value for the truncated joint model, compared to the joint and cure fraction models. The interpretation of this change is quite intuitive since the model is now not considering the existence of patients presenting no seizures pre-randomization. By not considering this case, the truncated density concentrates its mass towards positive values and considers a

lower heterogeneity between the individuals in the study.

Observe that while the pre-randomization parameter estimates remain essentially similar between the joint model and the joint model with a cure fraction, in the case of the zero-truncated model the effect of age in the patients has become more significant, as is the effect of the presence of Tonic-Clonic type of seizure only. Additionally, the treatment allocation covariate and its interaction with the EEG outcome appears to be of higher significance, relative to the other two models. Although the significance of the rest of the covariates do not suggest a difference of interpretation between the three models, the loglikelihood estimation shows that, in terms of the covariates considered, the joint model with cure fraction presents a better fit than that of the joint model, but the zero-truncated joint model proposes a better fit still. According to the size of the log-likelihood, we expect the zero-truncated model to adjust better than the original Cowling model, but in general its comparison to the Rogers model could be studied further. A possible next step of research could then consist of developing a zero truncated joint model considering a cure rate. In this particular case, the cure rate for the joint model with cure fraction is found to be

$$\hat{p} = \frac{\exp(\hat{\kappa})}{1+\exp(\hat{\kappa})} = \frac{\exp(0.004)}{1+\exp(0.004)} = 0.501,$$

which indicates that the estimate proportion of the patients in the population who do not experience a seizure after randomization, is of 50.10%. The corresponding standard error of the estimate κ, however, is substantially large compared to the estimated value, hence the cure fraction effect could be negligible.

For the three models, the EEG outcome appears to be non-significant due to the size of the corresponding standard error, as well as its interactions, in which only the interaction between EEG and patients presenting second generalized Tonic-Clonic seizures present a borderline significant decrease of the seizure rate compared to the reference population, which considers patients with no EEG abnormality and

partial seizures only. This borderline significance could be due, however, to the fact that the treatment allocation effect and its corresponding interactions tend to be significant. Additionally, the epilepsy types, both Tonic-Clonic only and Second generalized Tonic-Clonic, have higher estimated coefficients for the truncated joint model. This could be due to an influence of the epilepsy syndrome on the homogeneity level on the population. For this reason, in the following Table 4.2 we partition the epilepsy data by the syndrome of the patients, and fit the three models to each subset. A significant difference of the estimates of α between sub-sets indicates that a model considering a treatment effect on the homogeneity rate could be beneficial.

For the Table 4.2, we partition the epilepsy data by type of epilepsy syndrome, leaving us with three data subsets consisting of patients with Tonic-Clonic seizures only, patients with second generalized Tonic-Clonic seizures and finally individuals presenting partial seizures only. Observe that the population with Tonic-Clonic seizures only consists of 778 patients, followed by 453 patients with 2nd. Tonic-Clonic seizures, and finally we have 103 patients presenting partial seizures only. Under this comparison between models, the estimated age coefficient attains in general a higher value for both pre-randomization and post-randomization seizure recurrence rates for the truncated joint model, being deemed significant for patients with Tonic-Clonic and second generalized seizures, both for the recurrence rates before and after randomization. This behavior is not observed in the joint model and the joint model with cure fraction, where the age covariate seems to have a much smaller contribution to the seizure rate recurrence.

The clearest difference of homogeneity levels between models is shown for the population with Tonic-Clonic seizures only. Observe that the estimate of α under the truncated model is much greater than those under the Cowling and Rogers models. This could be an indication that, taking away the possibility of not having any seizures from the model, data set of 778 patients show a much smaller heterogeneity. The main difference between the three models for this subpopulation, is that for the

truncated joint model the treatment allocation seems to have little effect compared to the size of its standard deviation, indicating that the model does not consider immediate treatment as a significant factor for the change in seizure recurrence rate, as opposed to the joint and joint with cure fraction models. In contrast, for the truncated model the EEG abnormality becomes a much more significant indicator of the seizure recurrence post-randomization.

In the case of patients presenting second generalized Tonic-Clonic seizures, the truncated joint model continues to show the highest value for the frailty term α between models, and while the post-randomization factors seem to have a larger effect on the seizure rate in general, the most noticeable difference between the models is that the EEG factor and its interaction becomes quite significant, in contrast to the corresponding estimates given by the joint model with a cure fraction.

For the patients presenting partial seizures only, the frailty term appears to be the estimate that varies the least between models from the three subpopulations under consideration. It is also the population for whom the lowest values of α are estimated, consistently predicting a higher heterogeneity between individuals for the relatively small population of 103 patients. In the case of the truncated joint model, there appears to be a correlation between the heterogeneity estimated value and the significance of the age covariate both before and after randomization, but this could be more readily explained by the fact that, as observed in the bean-plots in Figure (2.2), patients with partial seizures range between a few months to 80 years old, with an empirical density showing a heavier tail than that observed for the other two subpopulations.

		Cowling	SE	Truncated	SE	Rogers	SE
	T-C only						
α		3.306	0.308	22.226	0.189	3.310	0.304
λ	Intercept	-5.265	0.040	-6.221	0.047	-5.257	0.040
	Age	-0.006	0.002	-0.022	0.002	-0.006	0.002
ψ	Intercept	-2.351	0.114	-1.831	0.125	-0.487	0.152
	Age	0.006	0.004	0.021	0.004	0.004	0.005
	Trt	-0.214	0.161	-0.083	0.167	-0.438	0.226
	EEG	0.409	0.162	0.762	0.155	-0.004	0.208
	Trt-EEG	-0.674	0.248	-1.114	0.252	-0.493	0.353
κ						0.0083	0.089
	-LogLikelihood	4038.215		3521.900		3939.972	
	2nd. Generalized						
α		1.676	0.143	14.686	0.248	1.619	0.137
λ	Intercept	-4.815	0.050	-5.203	0.039	-4.807	0.051
	Age	-0.005	0.003	-0.014	0.003	-0.005	0.003
ψ	Intercept	-3.014	0.164	-5.060	0.456	-1.581	0.242
	Age	0.005	0.005	0.019	0.007	-0.002	0.007
	Trt	-0.171	0.233	0.424	0.565	0.003	0.363
	EEG	0.969	0.229	2.740	0.487	1.313	0.304
	Trt-EEG	-1.029	0.338	-1.308	0.625	-1.298	0.474
κ						0.015	0.141
	-LogLikelihood	2512.854		2682.571		2443.513	
	Partials only						
α		1.049	0.158	1.419	0.383	1.027	0.154
λ	Intercept	-4.107	0.110	-4.126	0.105	-4.107	0.111
	Age	-0.001	0.004	-0.0007	0.004	-0.001	0.004
ψ	Intercept	-5.124	0.722	-10.751	5.198	-3.059	0.937
	Age	-0.012	0.012	-0.169	0.173	-0.087	0.028
	Trt	0.956	0.890	1.418	3.624	2.035	1.121
	EEG	-0.309	0.832	0.894	2.656	1.171	1.149
	Trt-EEG	-0.348	1.084	-1.929	4.032	-3.809	1.436
κ						-1.109	0.299
	-LogLikelihood	451.5491		487.697		436.117	

Table 4.2: Cowling, Truncated Cowling and Rogers models fitted to the epilepsy data partitioned by epilepsy syndrome. The first parameter estimates correspond to the population presenting Tonic-Clonic seizures only, the second set of parameter estimates correspond to the patients presenting Second Tonic-Clonic seizures, and the final set of estimated parameters correspond to the population presenting Partial seizures only.

		Cowling	SE	Truncated	SE	Rogers	SE
	T-C only						
α		3.306	0.154	20.179	0.091	3.310	0.152
λ	Intercept	-5.265	0.020	-6.191	0.023	-5.257	0.020
	Age	-0.006	0.001	-0.021	0.001	-0.006	0.001
ψ	Intercept	-2.351	0.057	-1.866	0.063	-0.487	0.076
	Age	0.006	0.002	0.020	0.002	0.004	0.002
	Trt	-0.214	0.081	-0.108	0.084	-0.438	0.113
	EEG	0.409	0.081	0.755	0.078	-0.004	0.104
	Trt-EEG	-0.674	0.124	-1.126	0.128	-0.493	0.176
κ						0.008	0.045
	-LogLikelihood	16152.86		14092.94		15759.89	
	2nd. Generalized						
α		1.675	0.072	13.774	0.124	1.619	0.069
λ	Intercept	-4.815	0.025	-5.139	0.019	-4.807	0.025
	Age	-0.005	0.001	-0.014	0.0012	-0.005	0.001
ψ	Intercept	-3.014	0.082	-5.234	0.241	-1.581	0.121
	Age	0.005	0.003	0.019	0.004	-0.002	0.003
	Trt	-0.170	0.117	0.348	0.303	0.002	0.181
	EEG	0.968	0.114	2.654	0.259	1.313	0.152
	Trt-EEG	-1.029	0.169	-1.342	0.338	-1.297	0.237
κ						0.015	0.071
	-LogLikelihood	10051.41		10793.09		9774.052	
	Partials only						
α		1.049	0.056	1.420	0.135	1.027	0.054
λ	Intercept	-4.107	0.039	-4.125	0.037	-4.108	0.039
	Age	-0.001	0.002	-0.0001	0.001	-0.001	0.002
ψ	Intercept	-5.124	0.255	-10.751	1.864	-3.081	0.336
	Age	-0.012	0.004	-0.167	0.062	-0.087	0.010
	Trt	0.957	0.315	1.418	1.309	2.062	0.399
	EEG	-0.308	0.294	0.894	0.963	1.195	0.410
	Trt-EEG	-0.349	0.383	-1.929	1.456	-3.839	0.510
κ						-1.106	0.106
	-LogLikelihood	3612.392		3902.721		3488.937	

Table 4.3: coefficient estimates when replicating the data four times for the T-C and 2nd T-C populations, and eight times for the Partials.

The truncated model for the partials only subpopulation differs from the other two models mainly in that it does not present the treatment allocation as a significant covariate for the contribution of ψ_i. It does produce, however, the lowest

loglikelihood maximized value of the three models for this and the second generalized seizure population, where the joint model with cure fraction consistently show the best fit. The Tonic-Clonic only subpopulation contrasts this finding by presenting the largest loglikelihood value under the truncated model. These differences in the corresponding fits, particularly in the case of partial seizures, could be due either to the small size of the subset under study, or to the fact that patients with partial seizures present much more frequent epileptic attacks than those patients with general seizures. This will effectively lead to very greater seizure count recordings, and in this case, the zero-truncated model does not produce a significant beneficial effect.

In an attempt to observe if the size of the subset only is the cause of the great discrepancy between models, we fit the three models to the partitioned data sets, where now each one has been replicated a number of times. The subsets of patients with Tonic-Clonic and Second Tonic-Clonic seizures have been replicated four times, whilst the subset of patients with partial seizures has been replicated eight times. The resulting estimates are shown in Table 4.3. The estimates for the population with partial seizures only shows that there is no change in the behaviour previously observed between models, leading us to believe that the size of the population is not the source of the discrepancy by itself.

4.1.5 Application of the Truncated Joint Model to the SANAD Data

Let us now remember that, in Section (2.2), we introduced the epilepsy database SANAD, which consisted of a study of two arms, A and B, dedicated to patients presenting partial or general seizures respectively. In this section we consider the fit of the joint model and the zero-truncated joint model for this data. The joint model with a cure fraction has also been considered for this study, however, as observed in the Kaplan-Meier plots shown in Figure (2.3), the existence of a cure rate is in

doubt, and the model therefore does not converge under any of the two arms in consideration.

In a preliminary analysis of the data it was observed that the number of seizures in the population ranged from 1 seizure to 30000, having a mean of 157.5 and the 95% quantile of 302.25 seizures. One approach to fit the data is by considering only the patients who presented less than 300 seizures pre-randomization. For this subset, we observe in Table 4.4 that a total 62 of cases are left out from the population, which consists of the 6.48% of the data. 38(5.17%) of these patients belong to Arm A, and the remaining 24 patients belong to Arm B. This leaves us with 94.82% of the population in Arm A, and 89.18% of the population from Arm B.

	General	Partial	Unclassified
Full data set	133	636	125
Arm A	8	624	64
Arm B	125	12	61

Table 4.4: Number of patients by syndrome and by Arm, with less than 300 seizures pre-randomization.

In the following Table 4.5 and Table 4.6, we summarize the covariates present for the subset of 624 patients in Arm A with partial seizures and no more than 300 seizures pre-randomization. The first table shows the discrete-valued covariates, while the second table displays a compendium of the quantiles for the continuous-valued covariates such as age at randomization, period and time to first seizure. Let us remember that the age is centered at 40 years of age.

For this reason, in Table 4.7 we show the model estimates both considering the entire data set, and considering the subset containing the data of patients with no more than 300 seizures. Observe that the values shown correspond to the estimates of the Gamma parameter α, the pre-randomization covariate estimates β_1 and the post-randomization covariate estimates β_2. In the first three columns we consider

Sex	Number	**Drug**	Number
Female	300	CBZ	146
Male	324	GBP	137
EEG		LTG	124
Unknown	50	OXC	81
No	302	TPM	136
Yes	272		

Table 4.5: Quantity of patients in Arm A with partial seizures only, categorized by sex, EEG outcome and Drug.

	Minimum	1st Quartile	Median	3rd Quartile	Maximum
(Age (in years)-40)/10	-3.49	-1.75	-0.33	1.26	4.31
Time (in days)	1	7.75	63.00	389.25	2414.00
Number of seizures	1	5	41	70	250
Period (in days)	182	203	516	2066	21366
Rates (in years)	0.103	3.577	7.473	19.890	461.60

Table 4.6: Summary of patients in Arm A with partial seizures only, categorized by Age at randomization, time of first seizure post-randomization, number of seizures, period in days and rates in years.

the model with the full data set, setting CBZ as the underlying baseline drug for the Full model and the Arm A model, and setting VPS as the underlying baseline drug for the model considering only Arm B. The same drug baselines are considered for the models Subset, Subset Arm A and Subset Arm B, in which the data has been restricted only to patients with no more than 300 observed seizures.

It is noticeable that, while the covariate coefficient estimates in general do not change significantly from the full data to the subset data for the pre-randomization parameter λ, a significant change is observed for the post-randomization covariate estimates of ψ, in all three categories (Full, Arm A and Arm B). In particular, we observe that for the estimates of ψ, the drug effect appears to be non-significant for the full data and the full data under Arm A only, whilst it appears to be significant for the full data under Arm B; this significance predictions, however, are reversed once the patients with more than 302 seizures are removed from the corresponding data sets, as can be observed by comparing their estimate magnitude and its stan-

	Full	Full Arm A	Full Arm B	Sub.	Sub. Arm A	Sub. Arm B
α	0.46 (0.02)	0.57 (0.02)	0.36 (0.03)	0.64 (0.03)	0.72 (0.03)	0.57 (0.05)
λ						
Interc.	-2.92 (0.23)	-4.05 (0.24)	-2.89 (0.52)	-3.86 (0.18)	-4.07 (0.21)	-4.19 (0.39)
General	1.75 (0.19)	-0.28 (0.53)	0.88 (0.27)	1.70 (0.17)	-0.24 (0.48)	1.02 (0.25)
Partial	1.02 (0.16)	1.97 (0.18)	1.67 (0.50)	1.43 (0.13)	1.55 (0.17)	1.80 (0.51)
Sex	-0.30 (0.10)	-0.22 (0.10)	-0.53 (0.25)	-0.27 (0.09)	-0.22 (0.09)	-0.37 (0.20)
Age	-0.26 (0.03)	-0.12 (0.03)	-0.76 (0.06)	-0.13 (0.02)	-0.08 (0.02)	-0.65 (0.08)
ψ						
Interc.	-2.83 (0.22)	-1.58 (0.24)	-3.04 (0.45)	-2.70 (0.22)	-2.43 (0.26)	-2.47 (0.43)
General	0.06 (0.16)	0.96 (0.52)	0.34 (0.27)	-1.48 (0.17)	0.90 (0.54)	-1.14 (0.25)
Partial	0.63 (0.14)	-0.77 (0.19)	-0.12 (0.41)	-0.33 (0.14)	-0.57 (0.19)	0.57 (0.46)
Sex	-0.32 (0.08)	-0.26 (0.09)	-0.17 (0.19)	0.20 (0.09)	0.17 (0.10)	0.36 (0.20)
Age	0.01 (0.02)	-0.05 (0.02)	0.28 (0.09)	-0.15 (0.02)	-0.17 (0.02)	-0.03 (0.10)
Drug	0.11 (0.12)	0.21 (0.12)	0.36 (0.24)	0.46 (0.12)	0.47 (0.12)	-0.21 (0.21)
-Loglik.	10220.15	7718.914	2346.95	9217.77	7155.122	1985.328

Table 4.7: Cowling's model estimates for the full population, for Arm A only, for Arm B only, and their corresponding versions with number of seizures$<$ 302.

dard deviation. From the values of α we can also observe that the estimated values are smaller for the full data set, giving and indication that the data corresponding to number of seizures larger than 302 confer a higher heterogeneity to the population.

We now consider fitting the joint model and the zero-truncated joint model to the two arms provided by the SANAD study, as a method of comparison of such models for the two epilepsy studies. From the log-likelihoods shown in Table 4.7 we observe that the model provides the best fits for the sub-populations corresponding to Arm A and Arm B when the patients present no more than 302 seizures, and thusly we proceed to compare the joint and the truncated joint models under these sub-populations.

For patients presenting partial seizures only and treated in Arm A, the pre-randomization coefficient estimates for λ (shown in Table 4.8) present almost no change in significance among the two models considered. For the post-randomization coefficients corresponding to ψ, however, the truncated joint model presents a higher significance for the age covariate, and even reverses the effect of the drugs Gabapentin (GBP) and Oxcarmazipan (OXC) on the seizure recurrence, relative to the reference population which is treated with Carbamazepine (CBZ). In contrast, the age

		Cowling	SE	Truncated	SE
α		0.68	0.03	0.86	0.05
λ	Intercept	-2.55	0.16	-2.54	0.10
	Sex	-0.27	0.10	-0.27	0.06
	Age	-0.08	0.03	-0.08	0.02
	Age^2	0.03	0.01	0.02	0.01
ψ	Intercept	-2.92	0.37	-1.74	0.33
	Sex	0.34	0.24	-0.08	0.21
	Age	0.09	0.07	-0.37	0.06
	Age^2	-0.10	0.02	-0.16	0.02
	LTG	1.70	0.54	0.78	0.46
	TPM	0.30	0.49	0.21	0.45
	GBP	-0.58	0.51	0.90	0.45
	OXC	1.61	0.58	-0.06	0.55
	Sex:LTG	-0.89	0.34	-0.09	0.29
	Sex:TPM	-0.01	0.32	-0.15	0.30
	Sex:GBP	1.28	0.32	0.23	0.31
	Sex:OXC	-0.87	0.37	0.19	0.34
	Age:LTG	-0.45	0.09	0.20	0.08
	Age:TPM	-0.23	0.09	0.16	0.09
	Age:GBP	-0.23	0.10	0.24	0.10
	Age:OXC	-0.13	0.11	0.08	0.10
	-Loglikelihood	6401.20		6514.43	

Table 4.8: Joint and zero-truncated joint model's estimates and their standard deviations, respectively, for the subpopulation in Arm A only that have experienced no more than 302 seizures.

covariate shows a diminished significance under the truncated joint model, when considering patients presenting general seizures in Arm B (shown in Table 4.9). The effects of the drugs Lamotrigine (LTG) and Topiramite (TPM) are predicted to have reverse effects on the seizure recurrence when compared to the effect of Valproate (VPS), when considered under the truncated joint model as opposed to the predictions under the joint model. Observe, however, that under both Arms the log-likehood values signal a better fit of the SANAD data under the joint model than the fit under the truncated model. Although there are reasonable conjectures about the reason of this discrepancy, mainly the small sample size in Arm B, and the inherent heterogeneity in Arm A due to partial seizures, a more throughout

study of goodness of fit is desirable as a future work project. As can be noticed in the following chapter, the goodness of fit studies are not performed for the SANAD study, since it is only intended here as an alternative comparison for the model predictions, but it does not constitute our main aim of study.

		Cowling	SD1	Truncated	SD
α		0.52	0.06	0.73	0.08
λ	Intercept	-4.53	0.45	-4.48	0.30
	Sex	-0.10	0.25	-0.08	0.16
	Age	-0.94	0.19	-0.84	0.11
	Age^2	0.06	0.07	0.07	0.04
ψ	Intercept	-2.48	1.03	-4.21	1.73
	Sex	-0.23	0.45	-0.23	0.68
	Ager	0.68	0.21	0.08	0.46
	$Ager^2$	0.28	0.07	0.17	0.11
	LTG	-1.07	1.30	1.92	2.06
	TPM	-2.44	1.33	2.83	1.95
	Sex:LTG	2.33	0.60	0.90	0.91
	Sex:TPM	0.84	0.60	-1.47	0.83
	Age:LTG	0.26	0.31	0.55	0.52
	Age:TPM	-0.07	0.35	0.29	0.52
-Loglikelihood		1286.95		1332.27	

Table 4.9: Joint and zero-truncated joint model's estimates and their standard deviations, respectively, for the subpopulation in Arm B only that have experienced no more than 302 seizures.

4.2 Frailty term depending on covariates

For all the previous models shown here, we have considered the frailty term α to be independent of the covariates. However, the change in the values of this parameter observed in Table 4.2 and in Table 4.3, there is evidence that the frailty term itself shows variation depending on the type of seizure or the EEG outcome of the patients. We propose, for each patient i, a joint model of the pre-randomization seizure count X_i and the post-randomization seizure first occurrences Y_i as proposed by Cowling, with the difference that, for each patient i in the population, the frailty term α_i will

be expressed in the form $\alpha_i = \exp(\beta_3 Z_{3i})$. Here Z_{3i} will represent the explanatory variables pre-randomization.

The likelihood, loglikelihood and derivative expressions with respect to β_1 and β_2 remain the same as the original joint model. The difference in notation will lie in the use of α_i instead of α, for each patient i. The corresponding first derivative of the log-likelihood with respect to β_3 is given by

$$\begin{aligned}\frac{\partial \ell(\alpha, \beta_1, \beta_2; x, y)}{\partial \beta_3} &= \sum_{i=1}^{n} \alpha_i z_3 \left\{ \sum_{j=0}^{x_i - 1} \left(\frac{1}{\alpha_i + j} \right) + \log(\alpha_i) + 1 + \frac{\delta_i}{x_i + \alpha_i} - \log(\zeta_i) \right. \\ &\qquad \left. - \frac{x_i + \alpha_i + \delta_i}{\zeta_i} \right\}.\end{aligned}$$

Observe that for this equation, we have used the fact that

$$\sum_{i=1}^{n} \{\log \Gamma(x_i + \alpha_i) - \log \Gamma(\alpha_i)\} = \sum_{i=1}^{n} \left\{ \sum_{j=0}^{x_i - 1} \log(\alpha_i + j) \right\}.$$

In the following Table 4.10, we show the fitted values for the covariates corresponding to each of the parameters λ_i, ψ_i and either α for the first two models or α_i for the last two models. From left to right, we observe the fitted values for the joint model, followed by the values for the zero-truncated joint model discussed before. The third model consists of the joint model which considers α_i to depend on the type of seizure, and the fourth model considers the joint model with the frailty term depending on the EEG outcome only.

	Cowling	S.E.	Trunc.	S.E.	α_{Type}	S.E.	α_{EEG}	S.E.
α	2.095	0.116	10.083	0.125				
Int					0.049	0.151	0.873	0.078
T-C					1.138	0.177		
2nd. T-C					0.470	0.173		
EEG							-0.278	0.112
λ								
Int	-4.131	0.085	-4.411	0.062	-4.116	0.109	-4.153	0.087
Age	-0.005	0.002	-0.012	0.001	-0.005	0.002	-0.005	0.002
T-C	-1.085	0.094	-1.335	0.076	-1.147	0.116	-1.068	0.096
2nd.T-C	-0.697	0.097	-0.738	0.073	-0.700	0.120	-0.686	0.098
ψ								
Int	-5.303	0.564	-7.219	1.214	-5.294	0.572	-5.335	0.568
Age	0.004	0.003	0.012	0.004	0.004	0.003	0.004	0.003
Trt	1.178	0.532	1.688	1.038	1.280	0.537	1.210	0.534
T-C	2.945	0.569	3.896	1.215	2.920	0.576	2.967	0.572
2nd.T-C	2.322	0.576	3.032	1.225	2.335	0.583	2.348	0.579
Trt*T-C	-1.366	0.532	-1.367	1.031	-1.447	0.536	-1.397	0.534
Trt*2nd.T-C	-1.446	0.542	-1.655	1.036	-1.566	0.547	-1.479	0.544
EEG	0.059	0.554	0.927	1.093	-0.029	0.560	0.086	0.555
Trt*EEG	-0.808	0.199	-1.246	0.270	-0.791	0.197	-0.826	0.200
EEG*T-C	0.410	0.556	0.380	1.089	0.486	0.561	0.417	0.557
EEG*2nd.T-C	0.820	0.566	1.292	1.100	0.885	0.572	0.816	0.567
-Loglikelihood	7031.64		6799.42		7004.65		7028.53	

Table 4.10: Cowling, Truncated Cowling, α dependent on seizure type and α dependent on EEG outcome models fitted to the epilepsy data.

For the joint model and the truncated joint model, their estimated α value denotes the level of heterogeneity estimated by the models, and as observed previously, the truncated joint model indicates a much larger homogeneity in the population than that observed in the joint model. For the models in which the frailty term is assumed to depend on covariates, we consider the reference population to have partial seizures and a normal EEG outcome, correspondingly. We obtain the estimated covariate coefficients for α shown in Table 4.10, and from them we obtain the corresponding frailty estimates displayed in Table 4.11. In the cases where we consider α to depend on the seizure type, the model predicts a value of $\hat{\alpha_{T-C}} = 3.277$ for patients presenting Tonic-Clonic seizures only, and a value of $\hat{\alpha}_{2nd.T-C} = 1.680$, contrasting with $\hat{\alpha}_{Partial} = 1.050$. This means that for this model, the group of patients

showing partial seizures are deemed the most heterogenetic, whilst the patients with Tonic-Clonic seizures only are the most homogenetic between populations. This is consistent with our knowledge that in patients diagnosed with partial seizures inherently present a larger variability of frequency and symptoms than those presenting general seizures. Analogously, the model predicts a larger heterogeneity in the population presenting an abnormal EEG ($\hat{\alpha}_{AbNor} = 1.813$) than the one that does not ($\hat{\alpha}_{Nor} = 2.394$).

Partial only	Tonic-Clonic	2nd. Tonic-Clonic	Normal EEG	Abnormal EEG
1.050	3.277	1.680	2.394	1.813

Table 4.11: Estimated values of α, first dependent on the type of epilepsy or secondly on the EEG outcome, resulting from Table 4.10.

Observe in Table 4.10 that although there are minor changes of the estimated covariates between the joint model and the models with dependent frailty terms, their values remain throughout similar, and the interpretations remain unchanged. For the zero truncated joint model, as observed in previous model comparisons, shows a greater age significance, whilst the treatment allocation's significance diminishes in contrast to the other three models. Although the covariate significance interpretation is very similar between the joint model and the joint models with dependent frailty terms, the difference in log-likelihood measures indicate that the dependency of α on covariates provides a better fit to the data than the joint model; however, the truncated joint model has an even better fit, which suggests that a zero-truncation of the models with dependent α could be a better approach to the epilepsy phenomenon. The computation of the corresponding α covariant-dependent truncated model, its likelihood and derivatives could be found from the formulas derived from the truncated joint model, but due to time restraints we leave this generalization open for further development and research.

4.3 Truncated Model with Cure Fraction

The following work shows the steps to calculate the likelihood for a zero-truncated joint model which considers a cure fraction p, or remission, from the first seizure after randomization. Although the code has already being developed, some specific parts of the log-likelihood have to be tuned for a proper optimization.

We consider from the original joint model with cure fraction, proposed by Rogers, the following distributions for the pre-randomization seizure count X and the first seizure post-randomization Y. Consider as before $\alpha > 0, \lambda > 0, \psi > 0$. Then we have the following distributions:

$$\begin{aligned} f_{Y|\nu}(y_i|\nu_i, \lambda_i, \psi_i, p_i) &= p_i \lambda_i \psi_i \nu_i \exp(-\lambda_i \psi_i \nu_i y_i), \\ S_{Y|\nu}(y_i|\nu_i, \lambda_i, \psi_i, p_i) &= 1 - p_i + p_i \exp(-\lambda_i \psi_i \nu_i y_i), \\ f_{X|\nu}(x_i|\nu_i, \lambda_i, u_i) &= \frac{(\lambda_i \nu_i u_i)^{x_i} \exp(-\lambda_i \nu_i u_i)}{x_i!}, \\ g_{\nu}(\nu_i|\alpha) &= \frac{\alpha^{\alpha} \nu_i^{\alpha-1} \exp(-\alpha \nu_i)}{\Gamma(\alpha)}, \end{aligned}$$

where

$$p_i = \frac{\exp(\kappa \omega_i)}{1 + \exp(\kappa \omega_i)}$$

is the corresponding cure rate for first seizure post-randomization.

In the following subsections we will proceed to show how the joint distribution of X_i and Y_i is obtained for each patient i in the population, depending on whether we consider the case in which the seizure time is observed or censored. Once we have corroborated obtaining the same joint distributions as the ones proposed by Cowling ([10]), we will obtain the zero-truncated joint distributions.

4.3.1 Zero Truncated Joint distribution with cure fraction, with Y_i Observed

Let us consider the case when the seizure time for individual i, Y_i is observed. Then the corresponding joint distribution X_i and Y_i has the form:

$$\begin{aligned}
f_{X_i,Y_i}(x_i, y_i \quad | \quad u_i, \lambda_i, \psi_i, \alpha, p_i) &= \int_0^\infty f_X|\nu_i(x_i|u_i, \lambda_i, \nu_i) f_Y|\nu_i(y_i|\lambda_i, \psi_i, \nu_i, p_i) g_\nu(\nu_i|\alpha)\partial\nu_i \\
&= \int_0^\infty \frac{(\lambda_i\nu_i u_i)^{x_i} e^{-\lambda_i\nu_i u_i}}{x_i!} p_i\lambda_i\psi_i\nu_i e^{-\lambda_i\psi_i\nu_i y_i} \frac{\alpha^\alpha \nu_i^{\alpha-1} e^{-\alpha\nu_i}}{\Gamma(\alpha)}\partial\nu_i \\
&= \frac{p_i\lambda_i^{x_i+1} u_i^{x_i}\psi_i\alpha^\alpha}{x_i!\Gamma(\alpha)} \int_0^\infty \nu_i^{x_i+\alpha} e^{-\nu_i(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)}\partial\nu_i \\
&= p_i \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i}{x_i!\Gamma(\alpha)(\lambda_i u_i + \lambda_i\psi_i y_i + \alpha)^{x_i+\alpha+1}} \\
&= p_i \frac{\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)^{x_i}\lambda_i\psi_i}{x_i!\Gamma(\alpha){\zeta_i}^{x_i+\alpha+1}},
\end{aligned}$$

where $\zeta_i = \lambda_i u_i + \lambda_i\psi_i y_i + \alpha$. From this joint density we obtain the corresponding zero-truncated density for $X_i > 0$ and Y_i, shown below.

$$\begin{aligned}
f_{X>0,Y}(x_i, y_i|u_i, \lambda_i, \psi_i, \alpha, p_i) &= \frac{p_i\Gamma(x_i+\alpha+1)\alpha^\alpha(\lambda_i u_i)_i^x\lambda_i\psi_i}{x_i!\Gamma(\alpha){\zeta_i}^{x_i+\alpha+1}} \Big/ \left[1 - p_i\left(\frac{\alpha}{\lambda_i u_i + \alpha}\right)^\alpha\right] \\
&= \frac{p_i\Gamma(x_i+\alpha+1)\alpha^\alpha\lambda_i\psi_i(\lambda_i u_i)^{x_i}(\lambda_i u_i+\alpha)^\alpha}{x_i!\Gamma(\alpha)\zeta_i^{x_i+\alpha+1}[(\lambda_i u_i+\alpha)^\alpha - p_i\alpha^\alpha]},
\end{aligned}$$

and if we denote by $\eta_i = \lambda_i u_i + \alpha$, the joint density takes the form

$$= \frac{p_i\Gamma(x_i+\alpha+1)\alpha^\alpha\lambda_i\psi_i(\lambda_i u_i)^{x_i}\eta_i^\alpha}{x_i!\Gamma(\alpha)\zeta_i^{x_i+\alpha+1}[\eta_i^\alpha - p_i\alpha^\alpha]}.$$

4.3.2 Zero Truncated Joint distribution with cure fraction, with Y_i Censored

If we now consider the case in which the seizure time Y_i is unobserved, we derive the joint density of X_i and Y_i and obtain the following equations.

$$
\begin{aligned}
f_{X_i,Y_i}(x_i, Y_i > y_i \quad | \quad & u_i, \lambda_i, \psi_i, \alpha, p_i) \\
= & \int_0^\infty f_X|\nu_i(x_i|u_i, \lambda_i, \nu_i) S_Y|\nu_i(y_i|\lambda_i, \psi_i, \nu_i, p_i) g_\nu(\nu_i|\alpha) \partial\nu_i \\
= & \int_0^\infty \frac{(\lambda_i \nu_i u_i)^{x_i} e^{-\lambda_i \nu_i u_i}}{x_i!} \left[1 - p_i + p_i e^{-\lambda_i \psi_i \nu_i y_i}\right] \frac{\alpha^\alpha \nu_i^{\alpha-1} \exp(-\alpha\nu_i)}{\Gamma(\alpha)} \partial\nu_i \\
= & \frac{(\lambda_i + \alpha)^{x_i} \alpha^\alpha}{x_i! \Gamma(\alpha)} \left[(1 - p_i) \int_0^\infty \nu_i^{x_i+\alpha-1} e^{-\nu_i(\lambda_i u_i + \alpha)} \partial\nu_i \right. \\
& \qquad \left. + p_i \int_0^\infty \nu_i^{x_i+\alpha-1} e^{-\nu_i(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)} \partial\nu_i \right] \\
= & p_i \frac{\Gamma(x_i + \alpha_i) \alpha^\alpha (\lambda_i u_i)^{x_i}}{x_i! (\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^{x_i+\alpha}} + (1 - p_i) \frac{\Gamma(x_i + \alpha) \alpha^\alpha (\lambda_i u_i)^{x_i}}{x_i! \Gamma(\alpha) (\lambda_i u_i + \alpha)^{x_i+\alpha}} \\
= & \frac{\Gamma(x_i + \alpha) \alpha^\alpha (\lambda_i u_i)^{x_i}}{x_i! \Gamma(\alpha)} \left[\frac{p_i}{(\lambda_i u_i + \lambda_i \psi_i y_i + \alpha)^{x_i+\alpha}} + \frac{1 - p_i}{(\lambda_i u_i + \alpha)^{x_i+\alpha}}\right] \\
= & \frac{\Gamma(x_i + \alpha) \alpha^\alpha (\lambda_i u_i)^{x_i}}{x_i! \Gamma(\alpha)} \left[\frac{p_i}{{\zeta_i}^{x_i+\alpha}} + \frac{1 - p_i}{{\eta_i}^{x_i+\alpha}}\right].
\end{aligned}
$$

In a similar way as before, we now proceed to find the corresponding joint density function for the zero-truncated model under the assumption that Y_i is unobserved.

$$
\begin{aligned}
f_{X_i>0,Y_i}(x_i, Y_i > y_i \quad | \quad & u_i, \lambda_i, \psi_i, \alpha, p_i) \\
= & \frac{\Gamma(x_i + \alpha) \alpha^\alpha (\lambda_i u_i)^{x_i}}{x_i! \Gamma(\alpha)} \left[\frac{p_i}{{\zeta_i}^{x_i+\alpha}} + \frac{1 - p_i}{{\eta_i}^{x_i+\alpha}}\right] \Big/ \left[1 - p_i \left(\frac{\alpha}{\lambda_i u_i + \alpha}\right)^\alpha\right] \\
= & \frac{\Gamma(x_i + \alpha) \alpha^\alpha (\lambda_i u_i)^{x_i} \eta_i^\alpha [p_i \eta_i^{x_i+\alpha} + (1 - p_i) \zeta_i^{x_i+\alpha}]}{x_i! \Gamma(\alpha) (\zeta_i \eta_i)^{x_i+\alpha} (\eta_i^\alpha - p_i \alpha^\alpha)}.
\end{aligned}
$$

4.3.3 The Full Log-Likelihood and Derivatives

The likelihood function is the product of the density functions over all patients. We will use the notation δ_i to denote whether their survival time has been observed or censored. As indicated in the section corresponding to the truncated joint model,

the parameters λ_i and ψ_i depend on the covariate vectors β_1 and β_2 respectively in the form $\lambda_i = \exp(z_{1i}\beta_1)$ and $\psi_i = \exp(z_{2i}\beta_2)$. If x and y are respectively the sets of seizure counts and first seizure times post-randomization for all n patients, we obtain the equation:

$$\begin{aligned}
\mathcal{L}(\alpha, \beta_1, \beta_2, \kappa; x, y) &= \prod_{i=1}^{n} \left[\frac{p_i \Gamma(x_i + \alpha + 1)\alpha^\alpha \lambda_i \psi_i (\lambda_i u_i)^{x_i} \eta_i^\alpha}{x_i! \Gamma(\alpha) \zeta_i^{x_i+\alpha+1} [\eta_i^\alpha - p_i \alpha^\alpha]} \right]^{\delta_i} \\
&\quad \times \left[\frac{\Gamma(x_i + \alpha)\alpha^\alpha (\lambda_i u_i)^{x_i} \eta_i^\alpha [p_i \eta_i^{x_i+\alpha} + (1-p_i)\zeta_i^{x_i+\alpha}]}{x_i! \Gamma(\alpha)(\zeta_i \eta_i)^{x_i+\alpha}(\eta_i^\alpha - p_i \alpha^\alpha)} \right]^{1-\delta_i} \\
&= \prod_{i=1}^{n} \frac{\Gamma(x_i + \alpha)\alpha^\alpha (\lambda_i u_i)^{x_i} \eta_i^\alpha}{x_i! \Gamma(\alpha) \zeta_i^{x_i+\alpha}(\eta_i^\alpha - p_i \alpha^\alpha)} \left[\frac{p_i (x_i + \alpha) \lambda_i \psi_i}{\zeta_i} \right]^{\delta_i} \\
&\quad \times \left[\frac{p_i \eta_i^{x_i+\alpha} + (1-p_i)\zeta_i^{x_i+\alpha}}{\eta_i^{x_i+\alpha}} \right]^{1-\delta_i}.
\end{aligned}$$

By applying the logarithm to the likelihood function we obtain the log-likelihood function.

$$\begin{aligned}
\ell(\alpha, \beta_1, \beta_2, \kappa; x, y) &= \log[\mathcal{L}(\alpha, \beta_1, \beta_2, \kappa; x, y)] \\
&= \sum_{i=1}^{n} \{ \log \Gamma(x_i + \alpha) - \log \Gamma(\alpha) - \log(x_i!) + \alpha \log(\alpha \eta_i) \\
&\quad + x_i \log(\lambda_i u_i) - (x_i + \alpha) \log(\zeta_i) - log(\eta_i^\alpha - p_i \alpha^\alpha) \\
&\quad \delta_i [\log(x_i + \alpha) + \log(\lambda_i \psi_i) + \log(p_i) - \log(\zeta_i)] \\
&\quad + (1 - \delta_i)[\log(p_i \eta_i^{x_i+\alpha} + (1-p_i)\zeta_i^{x_i+\alpha}) - (x_i + \alpha)\log(\eta_i)] \} ,
\end{aligned}$$

which, after some grouping of the terms, results in the final form:

$$\begin{aligned}
\ell(\alpha,\beta_1,\beta_2,\kappa;x,y) &= \sum_{i=1}^{n}\{\log\Gamma(x_i+\alpha)-\log\Gamma(\alpha)-\log(x_i!)+\alpha\log(\alpha)+x_i\log(u_i)\\
&\quad -(x_i+\alpha)\log(\zeta_i)-\log(\eta_i^{\alpha}-p_i\alpha^{\alpha})+[\alpha-\delta_i(x_i+\alpha)]\log(\eta_i)\\
&\quad +(x_i+\delta_i)\log(\lambda_i)-(x_i+\alpha)\log(\zeta_i)\\
&\quad +\delta_i[\log(x_i+\alpha)+\log(\psi_i)+\log(p_i)-\log(\zeta_i)]\\
&\quad (1-\delta_i)[\log(p_i\eta_i^{x_i+\alpha}+(1-p_i)\zeta_i^{x_i+\alpha})-(x_i+\alpha)\log(\eta_i)]\Big\}.
\end{aligned}$$

The first order derivatives can now be found for each of the parameters of interest:

$$\begin{aligned}
\frac{\partial\ell(\alpha,\beta_1,\beta_2,\kappa;x,y)}{\partial\alpha} &= \sum_{i=1}^{n}\Bigg\{\frac{\partial\log\Gamma(x_i+\alpha)-\partial\log\Gamma(\alpha)}{\partial\alpha}+\log(\alpha)+1+\log(\eta_i)+\frac{\alpha}{\eta_i}-\log(\zeta_i)\\
&\quad -\frac{\alpha}{\zeta_i}-\frac{\alpha\eta_i^{\alpha-1}+\eta_i^{\alpha}\log(\eta_i)-[\log(\alpha)+1]p_i\alpha^{\alpha}}{\eta_i^{\alpha}-p_i\alpha^{\alpha}}+\delta_i\left[\frac{1}{x_i+\alpha}-\frac{1}{\zeta_i}\right]\\
&\quad +(1-\delta_i)\Bigg[\frac{p_i\eta_i^{x_i+\alpha-1}[x_i+\alpha+\eta_i\log(\eta_i)]}{p_i\eta_i^{x_i+\alpha}+(1-p_i)\zeta_i^{x_i+\alpha}}\\
&\qquad +\frac{(1-p_i)\zeta_i^{x_i+\alpha-1}[x_i+\alpha-\zeta_i\log(\zeta_i)]}{p_i\eta_i^{x_i+\alpha}+(1-p_i)\zeta_i^{x_i+\alpha}}\Bigg]\Bigg\},\\
\frac{\partial\ell(\alpha,\beta_1,\beta_2,\kappa;x,y)}{\partial\beta_1} &= \sum_{i=1}^{n}z_1\lambda_i\Bigg\{\frac{\alpha u_i}{\eta_i}+\frac{x_i}{\lambda_i}-\frac{(x_i+\alpha)(u_i+\psi_iy_i)}{\zeta_i}-\frac{\alpha\eta_i^{\alpha-1}u_i}{\eta_i^{\alpha}-p_i\alpha^{\alpha}}\\
&\quad +\delta_i\left[\frac{1}{\lambda_i}-\frac{u_i+\psi_iy_i}{\zeta_i}\right]\\
&\quad +(1-\delta_i)\Bigg[\frac{p_i(x_i+\alpha)\eta_i^{x_i+\alpha-1}u_i+(1-p_i)(x_i+\alpha)\zeta_i^{x_i+\alpha-1}(u_i+\psi_iy_i)}{p_i\eta_i^{x_i+\alpha}+(1-p_i)\zeta_i^{x-i+\alpha}}\\
&\qquad -\frac{(x_i+\alpha)u_i}{\eta_i}\Bigg]\Bigg\},\\
\frac{\partial\ell(\alpha,\beta_1,\beta_2,\kappa;x,y)}{\partial\beta_2} &= \sum_{i=1}^{n}z_2\psi_i\Bigg\{-\frac{(x_i+\alpha)\lambda_iy_i}{\zeta_i}+\delta_i\left[\frac{1}{\psi_i}-\frac{\lambda_iy_i}{\zeta_i}\right]\\
&\quad +(1-\delta_i)\left[\frac{(1-p_i)(x_i+\alpha)\zeta_i^{x_i+\alpha-1}\lambda_iy_i}{p_i\eta_i^{x_i+\alpha}+(1-p_i)\zeta_i^{x_i+\alpha}}\right]\Bigg\},\\
\frac{\partial\ell(\alpha,\beta_1,\beta_2,\kappa;x,y)}{\partial\kappa} &= \sum_{i=1}^{n}z_3p_i\omega_i\left\{-\frac{\alpha^{\alpha}}{\eta_i^{\alpha}-p_i\alpha^{\alpha}}+\frac{\delta_i}{p_i}+(1-\delta_i)\frac{\eta_i^{x_i+\alpha}-\zeta_i^{x_i+\alpha}}{p_i\eta_i^{x_i+\alpha}+(1-p_i)\zeta_i^{x_i+\alpha}}\right\}.
\end{aligned}$$

Let us remember that

$$p_i = \frac{\exp(\kappa z_3)}{1+\exp(\kappa z_3)},$$

hence, the derivative of this quantity with respect to κ is of the form

$$\begin{aligned}\frac{\partial p_i}{\partial \kappa} &= \frac{z_3 e^{\kappa z_3}(1+e^{\kappa z_3}) - z_3(e^{\kappa z_3})^2}{(1-e^{\kappa z_3})^2} \\ &= \frac{z_3 e^{\kappa z_3}(1+e^{\kappa z_3}-e^{\kappa z_3})}{(1+e^{\kappa z_3})^2} = p_i \frac{z_3}{1+e^{\kappa z_3}} = z_3 p_i(1-p_i).\end{aligned}$$

Through these equations we can find numerically the second-order derivatives, from which the standard deviation for the corresponding covariates under consideration are calculated. In Table 4.12 we show the corresponding maximum likelihood estimates of interest, contrasting the estimates between the zero-truncated joint model with a cure fraction (truncated Rogers model), the joint model with and without cure fraction (Cowling and Rogers models respectively) and the zero-truncated joint model (truncated Cowling model). It can be observed that estimate of α in truncated models indicates a higher homogeneity in the population than the non-truncated models, and furthermore, the post-randomization parameter standard errors are shown to be greater for the former than the latter models.

The truncated Rogers model presents no significant difference in the parameter estimates when compared to its non-cure fraction counterpart, the truncated Cowling model. However, when compared to the non-truncated Rogers model, the cure fraction parameter estimate, κ shows a difference from its truncated version, but it is not a significant difference, since the standard deviation is large enough to deem such differences unimportant. Indeed, when recalling that the cure fraction is

of the form

$$p = \frac{\exp(\kappa z)}{1 + \exp(\kappa z)},$$

we obtain that $p = 0.501$ for the Rogers model, and $p = 0.496$ for the zero-truncated Rogers model, differing in an additional 0.5% proportion of the population predicted to attain remission under the non-truncated Rogers model.

The log-likelihood value shows that the zero-truncated model with cure fraction does not seem to provide an improvement over the other three models. However, further investigation must be made about the properties of this model under extreme seizure count values, and other possible particularities. Due to the complexity of some of the expressions in the log-likelihood and its derivatives, when the number of seizures pre-randomization is high, the terms containing this variable tend to large or very small numbers which can cause problems during the optimization process. Further investigation and work for this model are left as a future work of interest.

		Cowling	SD	Trunc. Cowling	SD	Rogers	SD	Trunc. Rogers	SD
α		2.095	0.116	10.083	0.125	2.060	0.113	7.997	0.003
λ	Intercept	-4.131	0.085	-4.411	0.062	-4.129	0.085	-4.342	0.058
	Age	-0.005	0.002	-0.012	0.001	-0.005	0.002	-0.008	0.001
	T-C	-1.085	0.094	-1.335	0.076	-1.075	0.095	-1.227	0.065
	2T-C	-0.697	0.097	-0.738	0.073	-0.690	0.098	-0.796	0.066
ψ	Intercept	-5.303	0.564	-7.219	1.214	-4.045	0.615	-6.975	1.439
	Age	0.004	0.003	0.012	0.004	0.0001	0.004	0.011	0.005
	Trt	1.178	0.532	1.688	1.038	0.702	0.708	1.788	1.235
	T-C	2.945	0.569	3.896	1.215	3.513	0.626	4.044	1.442
	2T-C	2.322	0.576	3.032	1.225	2.565	0.638	3.120	1.457
	Trt*T-C	-1.366	0.532	-1.367	1.031	-1.009	0.715	-1.309	1.226
	Trt*2T-C	-1.446	0.542	-1.655	1.036	-0.942	0.731	-1.619	1.243
	EEG	0.059	0.554	0.927	1.093	0.652	0.712	1.024	1.292
	Trt*EEG	-0.808	0.199	-1.246	0.270	-0.860	0.281	-1.212	0.347
	EEG*T-C	0.409	0.556	0.380	1.089	-0.514	0.720	0.434	1.285
	EEG*2T-C	0.820	0.566	1.292	1.100	0.480	0.737	1.330	1.303
κ						0.004	0.076	-0.015	0.026
	-Loglike.	7031.64		6799.42		6861.38		8020.69	

Table 4.12: Maximum likelihood parameter estimates for the Cowling, Rogers and Truncated joint model with and without cure fraction (Truncated Cowling and Truncated Rogers models).

Chapter 5

Goodness of Fit Between Cox and Joint Models

In the following chapter we implement and explore the theory of residual analysis to four survival models, among other goodness of fit comparisons. We apply such theory and a few other model-checking techniques specifically to the epilepsy MESS data set.

When examining the adequacy of a model, the use and analysis of residuals has the potential for discovering outlying anomalous observations, influential observations or unexpected patterns. In the following section, we proceed to introduce the types of residuals that are commonly considered for survival analysis. The types of residuals are defined in terms of the estimated survival function at every event time t_i, and some of their most relevant properties are discussed. In the second section, we propose four survival models for the MESS data, and mainly four types of residuals are shown for each one of them. We then proceed to identify the outliers marked by the proportional hazards models, and we compare the model's fitted values when considering the original data set and the data set without the outliers.

In the third section we produce a simulation according to the assumptions from the Cowling model, in an attempt to investigate how the residuals would

behave under known conditions. Finally, in the last section we present the plots of Kaplan-Meier survival curves contrasting with the survival functions under the joint and joint with cure fraction models, taking the maximum likelihood estimated parameters under the MESS data set.

5.1 Introduction to Residual Analysis

The *Cox-Snell residuals* are commonly used in survival analysis, and a good discussion can be found in the book by Collet, 1994[7]. The genesis of such residuals comes from the principle that, if we denote by $S_i(t)$ the survival function for individual i at time t, then, by performing a change of variable, the random variable $Y = -\log S(T)$ is found to be distributed as an exponential distribution with mean one. Note that the variable is distributed as an exponential, irrespectively of the form of the survival function, which in particular means that this property holds for both proportional hazards models, accelerated life models or any other survival model. Furthermore, if the model fitted to the observed data is satisfactory, then the estimated survival function of individual i at time t_i, $\hat{S}_i(t_i)$, will approximate the real survival function $S_i(t_i)$, and the two functions will share similar properties. It will follow that, given that $-\log S_i(T_i)$ will be distributed as an $Exp(1)$, then $-\log \hat{S}_i(T_i)$, which is defined as the Cox-Snell residual and is denoted by r_{C_i}, will also be distributed thusly.

The Cox-Snell residuals have several properties which make them differ from the residuals used for linear regression models. The distribution of the residuals will not be symmetrically distributed around zero, as in fact they cannot be negative-valued and should have a highly skewed distribution. Since the residuals, under the right model, are distributed as an $Exp(1)$, they must show a unit mean and variance. A property of particular interest to us from the Cox-Snell residuals, is that if an individual has a right-hand censored survival time, then the corresponding value of

the residual will also be right-censored. This will mean that, if we have estimated the survival function for each patient in the study, we can propose a Cox-Snell residual study, where we will have that the censored survival times will in turn produce censored residuals. This property is of particular interest to us, since it provides a possible residual study for our epilepsy data, where we encounter different types of censoring throughout the follow up. How distinguishable would the censored data be, however, from the uncensored seizure times? Let us then consider the modified Cox-Snell residuals.

There is a refinement of the Cox-Snell residuals that has been proposed when in presence of right-censoring in the population. Suppose there is a censored survival time t_i^* for individual i, and that the true, but unknown, survival time for such individual is t_i. Then we will have that $t_i > t_i^*$, and hence, the cumulative hazard function evaluated at t_i must be greater than that evaluated at t_i^*, resulting in the Cox-Snell residual of t_i^* ($r_{C_i} = \hat{H}_i(t_i^*) = -\log \hat{S}_i(t_i^*)$) being underweighted. The *modified Cox-Snell residuals* aim to solve this problem by adding a positive quantity Δ to the original residuals, so that they are defined as

$$r'_{C_i} = \Delta\delta_i + r_{C_i} = \begin{cases} r_{C_i} & \text{for uncensored observations,} \\ r_{C_i} + \Delta & \text{for censored observations,} \end{cases}$$

where δ_i is the censoring indicator for an individual i, meaning that it will take the value zero if the observed survival time is censored, and one if it is uncensored. Observe that, by the lack of memory property of the exponential distribution, and since r_{C_i} is a unit exponential distributed, then the since the residuals are distributed as exponentials with mean one, then the excess residual Δ will also be distributed as a unit exponential. This will imply that the mean will be one, and we conclude

that $\Delta = 1$, leading to the modified Cox-Snell residuals

$$r'_{C_i} = \delta_i + r_{C_i} = \begin{cases} r_{C_i} & \text{for uncensored observations,} \\ r_{C_i} + 1 & \text{for censored observations.} \end{cases}$$

This will imply that all censored individuals will produce residuals with values greater than one. It has been argued, however, by Crowley and Hu[14], that considering $\Delta = 1$ results in overweighting the residuals, and propose to use, instead of the mean of the Exp(1), its median. This will result in considering $\Delta = \log(2) = 0.693$; which will result in the values of the residuals for the censored observations being greater than Δ. For our particular epilepsy model, this property would be useful, since it would allow us to observe the influence of the censored seizure times on the model.

Suppose a transformation is used on the Cox-Snell residuals, relocating the mean to zero when the observations are uncensored, and then multiplying by minus one. Then we would obtain the *martingale residuals* (also known as the *Lagakos residuals*), which are defined as

$$r_{M_i} = \delta_i - r_{C_i}.$$

These residuals are originally derived from probabilistic theory denoted as martingale methods, but such procedures are not derived here. This type of residual has some similarities to the residuals for generalized linear models, such as the residuals being centered at zero. However, they differ in that the residuals values range from $-\infty$ to 1, and the residual distribution is asymmetric around zero. However, there is a way of making the martingale residuals symmetric about zero when the model is satisfactory. Such a transformation derives to what is known as the *deviance residuals*. As the name suggests, these residuals arise from computing the deviance, which is a statistic designed to measure how much does a saturated model deviate

from a new proposed model. Such statistic is given by

$$D = -2\left\{\log \hat{L}_c - \log \hat{L}_f\right\},$$

where $\hat{L}_c$ is the maximized likelihood function under the new model, and $\hat{L}_f$ is the maximized likelihood function under the full model. The deviance residuals are then found to be of the form

$$r_{D_i} = sgn(r_{M_i})\left[-2\left\{r_{M_i} + \delta_i \log(\delta_i - r_{M_i})\right\}\right]^{1/2},$$

where $sgn(x)$ is the sign of x, and r_{M_i} is the martingale residual for the i^{th} patient. Although the deviance residuals are now symmetric around zero, their sum is not necessarily zero.

5.2 Residual Study for Epilepsy Data

A residual study is performed for four models. The first two models consist of a Cox proportional hazards models, but each one considers a different set of covariates to explain the times to seizure recurrence. For the model proposed by Kim et al.[27] the covariates considered are neurological disorder, EEG outcome and the logarithm of the number of seizures per patient, with the model stratifying by treatment allocation (immediate or deferred). For the model we denominate as the "Cox model", instead we consider the covariates treatment allocation (immediate or deferred), EEG outcome, logarithm of the number of seizures per patient, age at randomization and type of epilepsy, along with interactions. The Cowling or joint model is a Poisson process with over dispersion and a Gamma mixing distribution. For this model we consider the covariates EEG outcome, age at randomization, type of epilepsy, treatment allocation and their second order interactions. The final model is named the Rogers or joint with cure fraction model, and consists of the

Cowling model with a generalization of considering a cure rate, which represents the proportion of patients that will not present a seizure recurrence after randomization.

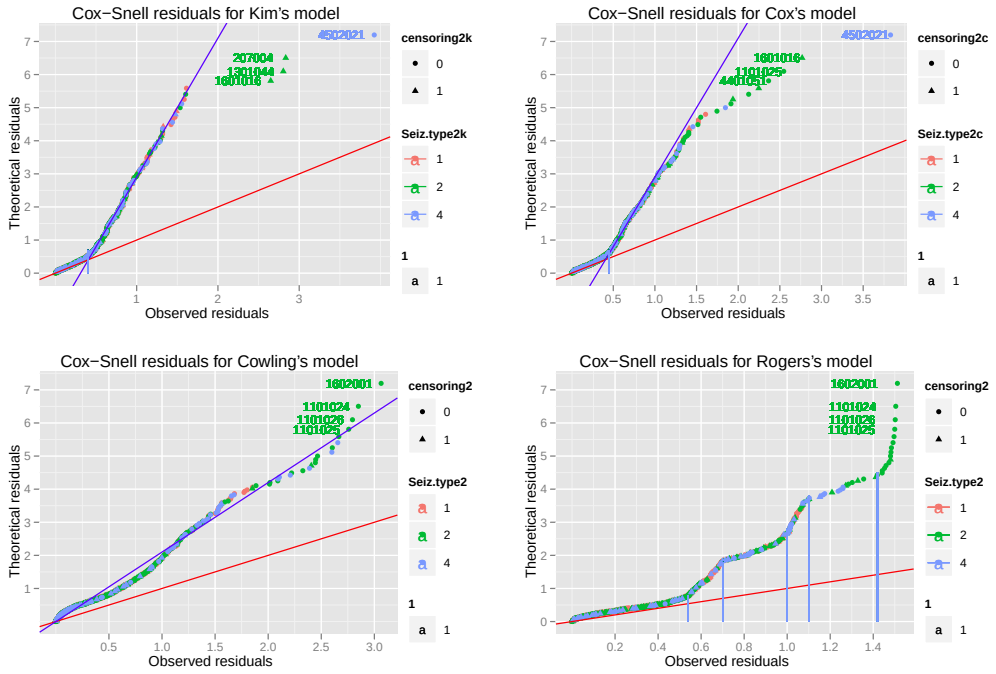


Figure 5.1: Q-Q plots for Cox-Snell residuals for Kim, Cox, Cowling and Rogers models.

In Figure 5.1 we observe the Cox-Snell residuals for each of the models. The ordered residuals are plotted against the theoretical values of an exponential distribution with unit mean. The straight red line through the origin and unit slope represents the ideal path of the residuals if the model is correct. Circles represent residuals obtained from observed times, while triangles correspond to the residuals from censored times. The three colors represent the three epilepsy types observed. Tonic-Clonic only (T-C), Second generalized Tonic-Clonic (2nd.T-C) and Partial seizures are colored red, green and blue respectively. Observe the plot corresponding to the Kim model shows that, for early seizures, about 40.5% of the data seems well fitted by the model. For the remaining observations, the straight blue line shown in the figure seems to fit better except for a few outliers. The blue line has an intercept

of -1.3 and a slope of 4.2, which means that for the considered model, the Cox-Snell residuals r_C have the form $r_C = -1.3 + 4.2Y$, where $Y \sim Exp(1)$. Under this new transformation, we identify four possible outliers, presented in Table 5.2. In order to compare the number of seizures, period of observation, rate and times to first seizures post randomization of these patients, we show in Table 5.1 the quartiles, mean, median, maximum and minimum of such variables. Considering that the mean number of seizures in the population is 2.49 and the third quartile is 2, the patients corresponding to the possible outliers experienced a much greater number of seizures. Observe that the patient with outlier identification number "1" has the greatest rate value in the population, due to the high number of seizures (65) in a short period of time (182 days).

For the Cox model, the Cox-Snell residuals, also shown in Figure 5.1, there is approximately a 46.7% of the population that is well adjusted by the model. Once again, the straight line with intercept -1.3 and a slope 4.2 appears to be a better fit for the remaining proportion of the population, and it can be seen in the plot as the blue straight line. Observe in Table 5.2 that once again there are three patients with 2nd.T-C seizures and one patient with partial seizures with the larger values of the residuals, and which are candidates for possible outliers. Two of these four residuals coincide with the possible outliers in the Kim model. Mainly, the residuals corresponding to patients numbered as "7" and "1". Although the four patients with greatest residual values in the Kim model range in ages from 6 to 40 years old, the corresponding patients for the Cox model consist of young people from 6 to 27 years old, and all of them present an abnormal EEG outcome. Although their number of seizures is higher than the mean number of seizures in the population, their seizure rates are not much greater than expected. Except, of course, for the patient numbered as "1".

An important characteristic that must be noticed for the residuals corresponding to these two models, is that all six outliers found correspond to female

	Minimum	1st. Quartile	Median	Mean	3rd. Quartile	Maximum
Num.Seiz	1	1	1	2.49	2	130
Times	1	109.2	724	949	1590	3161
Period	180	180	180	530	190	21000
Rate	0.0002	0.005	0.005	0.008	0.007	0.357

Table 5.1: Minimum, maximum, mean, and quartiles for number of seizures, times of first seizures post-randomization, period and rate.

patients; a phenomena that is observed again for the residuals corresponding to the Cowling and Rogers models, even when the outliers are not the same as the ones already discussed.

	Outlier								
	1	2	3	4	5	6	7	8	9
Model	K,Co	C,R	C,R	C,R	K	Co	K,Co	All	K
Sex	Fem	Fem	Fem	Fem	Fem	Fem	Fem	Fem	Fem
Age	6	15	15	15	17	18	25	28	40
Type	Partial	2.T-C	2.T-C	2.T-C	2.T-C	2.T-C	2.T-C	2.T-C	2.T-C
Trt	Imm	Imm	Imm	Imm	Def	Imm	Def	Imm	Imm
EEG	1	1	1	1	1	1	1	1	0
No.Seiz	65	1	3	1	31	6	104	5	100
Period	182	182	289	182	182	210	9300	769	9825
Time	1518	2497	2591	2978	1069	1440	774	2534	885
Cens	0	0	0	0	1	0	1	0	1
Rate	0.357	0.005	0.010	0.005	0.170	0.029	0.011	0.007	0.010
Centre	4502	1101	1101	1602	1301	4401	1601	1101	207

Table 5.2: Outliers and residuals with highest values corresponding to Kim (K), Cox (Co), Cowling (C) and Rogers (R) models. here 2.T-C represents the secondary generalized Tonic-Clonic seizures.

In the corresponding Cox-Snell residual plot in Figure 5.1 for the Cowling model, a different trend of the residuals can be noted. In this case the residuals do not appear to be well fitted from the beginning, but rather another straight line with intercept 0 and slope 2.1, presented in blue, seems to fit the data better. The residuals appear to oscillate around the new fitted line, and it could be a suggestion that a transformation could be used, such as a transformation of fourth power, or a sinusoidal transformation. In this case we also mark the four residuals with the greatest values, although under the new fit these do not appear to be outliers.

When observing the Cox-Snell residuals for the Rogers model in the same figure, it is noticeable that approximately the 54.8% of the population seem to be well fitted. For the remaining proportion of the residuals, there appears to be five sub-populations that could be depending on a covariate not yet considered in the model. Observe that there does not appear to be a pattern of type of epilepsy except for the residuals with the highest values, all of which present partial seizures only.

The four residuals with highest values are the same for both the Cowling and the Rogers model, and they are shown in Table 5.2. Observe however that all four patients corresponding to these residuals are female, allocated to immediate treatment, present abnormal EEG and 2nd.T-C seizures and are recorded as not censored. Additionally, three of these patients are aged 15 years old at the time of randomization. Another subset of three outliers belong to the same clinical centre, and their ID numbers are consecutive. This supports earlier suspicions that there is a centre factor that could be convenient to consider in the models. Observe that throughout the study 83 centres participated, and the number of patients they treated ranged from 1 to 238 individuals. In fact, for Outlier No.3, this patient was the only contribution by the centre to the study. As a part of future work, we are interested in analyzing further the relationship of the centre covariate to the possible biases of the seizure time reports.

When considering the joint model with a dependent heterogeneity parameter α on covariates, in this case, on the type of epilepsy, the corresponding Q-Q plot for the Cox-Snell residuals is shown in Figure 5.2. Once again the fitted straight line, as considered for the joint shown in Figure 5.1, is considered to have intercept 0 and slope 2.1, and is shown in blue. Observe that the fundamental contrast to the joint model with constant α is that the residuals for late seizure occurrences do not appear to oscillate, but rather remain close to the regression line. Apart from this variation, the general behaviour of the residuals for the rest of the seizure times remains similar to the joint model with fixed α. The observed highest valued

residuals remain the same and the four possible outliers under this model coincide with those obtained from the joint and joint with cure fraction models. This could be an indication that α depending on covariates provides a stabilizing effect on the model fitting.

Figure 5.2: Q-Q plot for the Cox-Snell residuals for the joint model with α dependent on the type of epilepsy.

In Figure 5.3, the martingale residuals are shown for the four models. Let us remember that the martingale residuals are centered at 0, taking values from -1 to infinity, with the residual density not being symmetrical around 0 and compressing all censored residuals from -1 to 0. The first two index plots are consistent in showing the individual with outlier number "1" (ID number "4502021") as an outlier for both the Kim and Cox models. The martingale residuals for the Cowling model presents less obvious outliers, but there seems to be a pattern in the data, indicating once again that there might be a variable that should be considered in the model. For the Rogers model, the index plot shows a more symmetrical behaviour around 0,

but the binary nature of the covariates considered in the model make more apparent the subsets of possible values that the residuals can take. For the Rogers model, there is no clear indication of an outlier.

The martingale residuals are transformed to obtain the deviance residuals, with the aim of obtaining a symmetric density of these new residuals around 0. The deviance is a statistic designed to measure how much does a saturated model deviate from a new proposed model. Although the deviance residuals are now symmetric around zero, their sum is not necessarily zero.

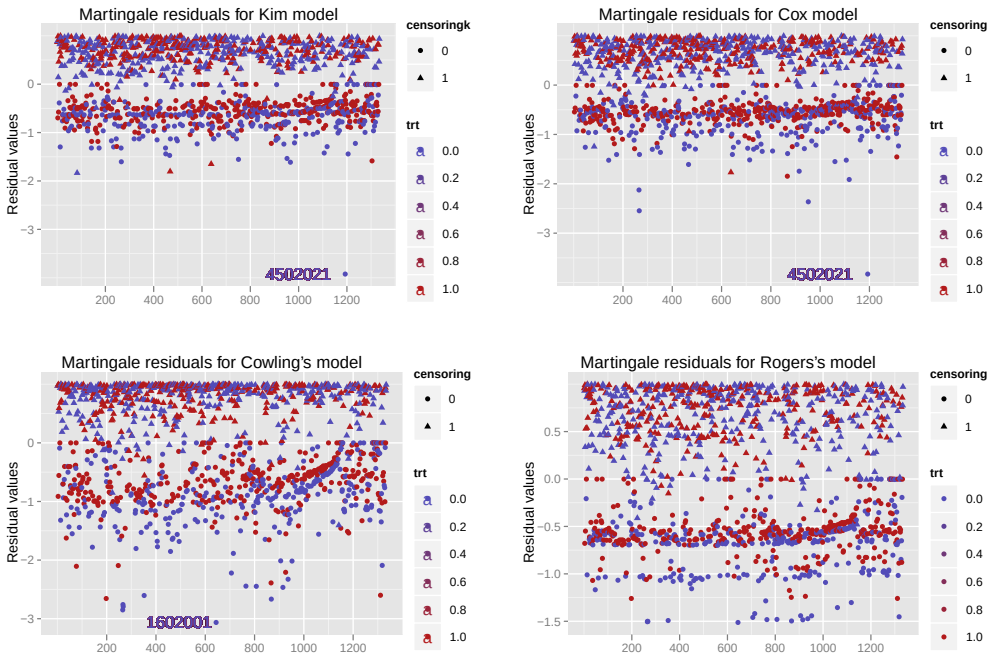


Figure 5.3: Martingale residuals for Kim, Cox, Cowling and Rogers models.

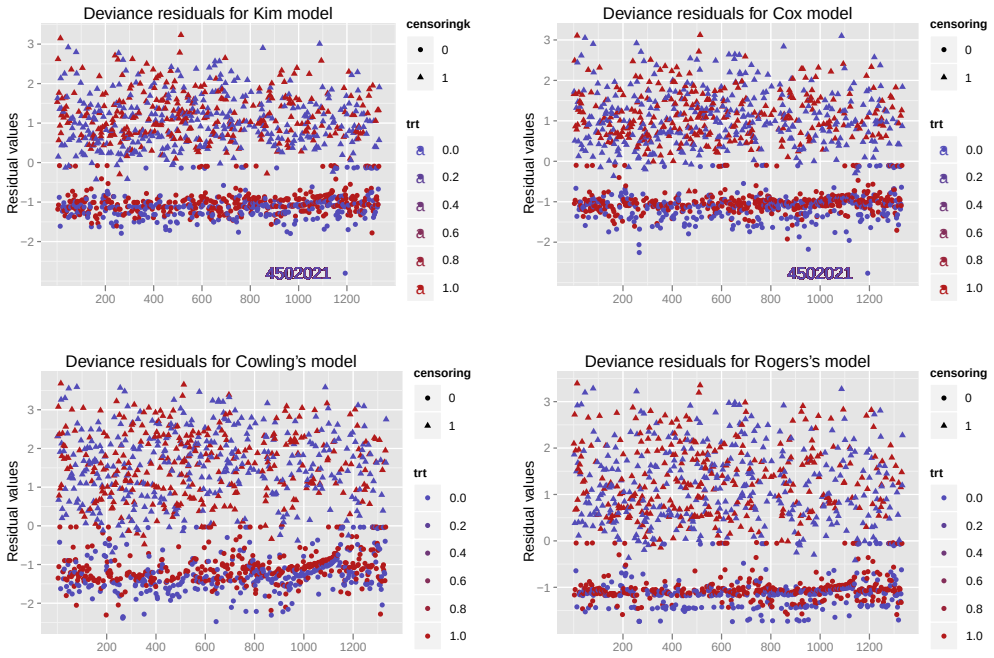


Figure 5.4: Deviance residuals for Kim, Cox, Cowling and Rogers models.

In Figure 5.4, we observe that the deviance residuals do not appear to be symmetrical around 0 for any of the four models. The index plots for the Kim and Cox models support the finding of the outlier previously observed, and the Cowling and Rogers models do not present outliers. The observed data is noticeably less sparse than the censored data, and for the four models the observed residuals are centered at -1, suggesting a deviation of such magnitude.

For neither the martingale nor the deviance residuals appears to be a trend for treatment allocation, as the density of blue-coloured, immediate-allocated patients is about the same as the density for the red-coloured, delayed-allocated patients. Furthermore, since our only continuous covariate is the age at randomization for these models, we plot the Cox-Snell residuals against this variable. In Figure 5.5 these plots show that there is no perceptible trend for the residuals of any of the models, although the Rogers model presents once more the effect of considering

binary covariates.

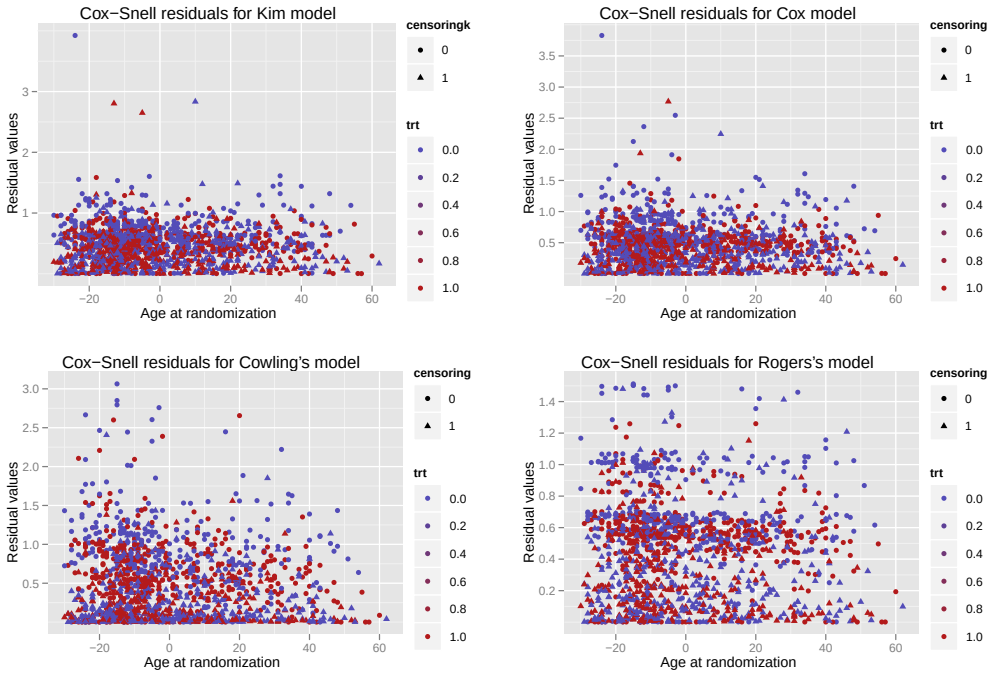


Figure 5.5: Cox-Snell residuals for Kim, Cox, Cowling and Rogers models, plotted against the age at randomization.

In Figure 5.6 we contrast the Cox-Snell residuals between models, in an attempt to observe if the models are consistent from one to another. The first plot shows the residuals corresponding to the Kim model against the corresponding residuals for the Cox model. We observe that, although the variation increases with time, the residuals are consistently centered around the line passing through 0 with unit slope. A comparison between Kim's residuals against Rogers' residuals, we again have that the residuals are dispersed around the equality line, but they appear skewed upwards, as roughly happens for the residuals of the Cox model against the Rogers model. For the plots of the Cowling model against Kim and Cox models, in both situations there seems to be a translated and re-scaled logarithmic transformation of the Cowling residuals. For the comparison between the Rogers and Cowling model, we observe that first seems to be an exponential transformation

of the second model, and that both models categorize the residuals in the same treatment-dependent binary combinations.

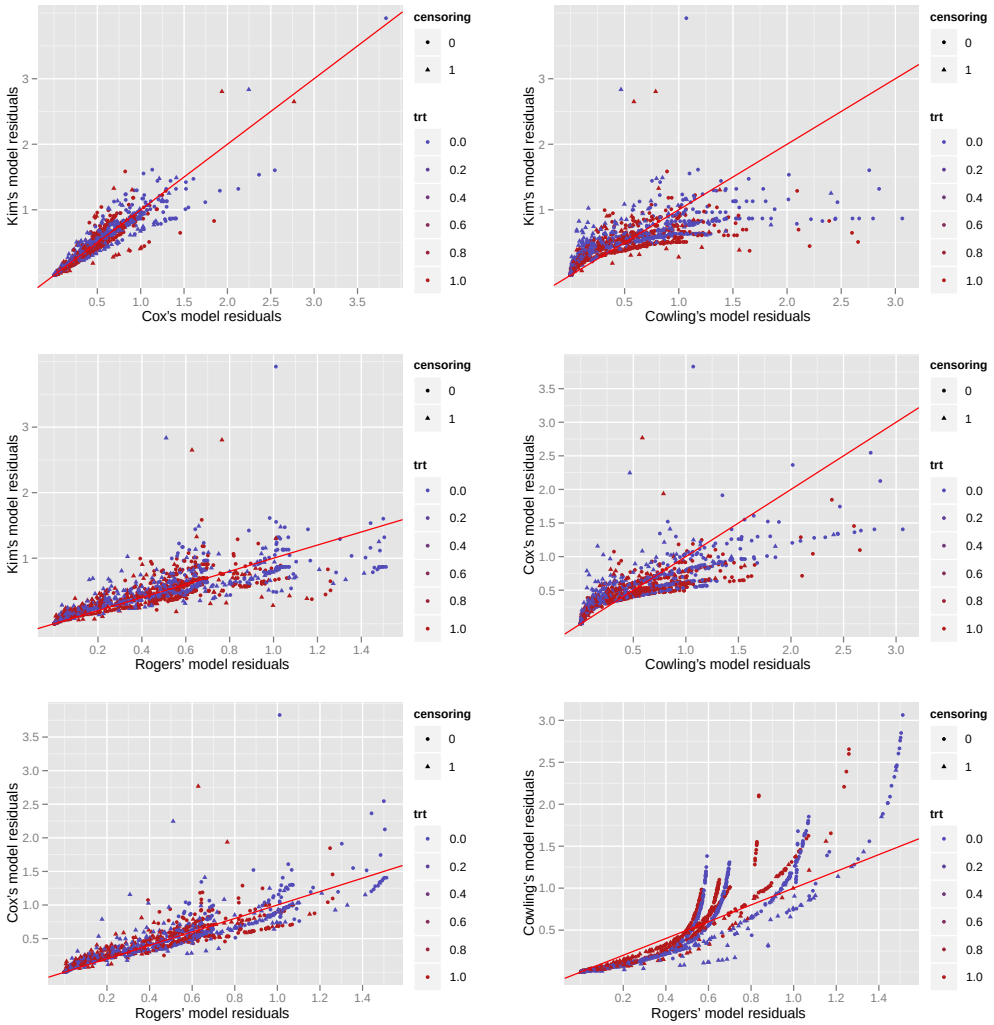


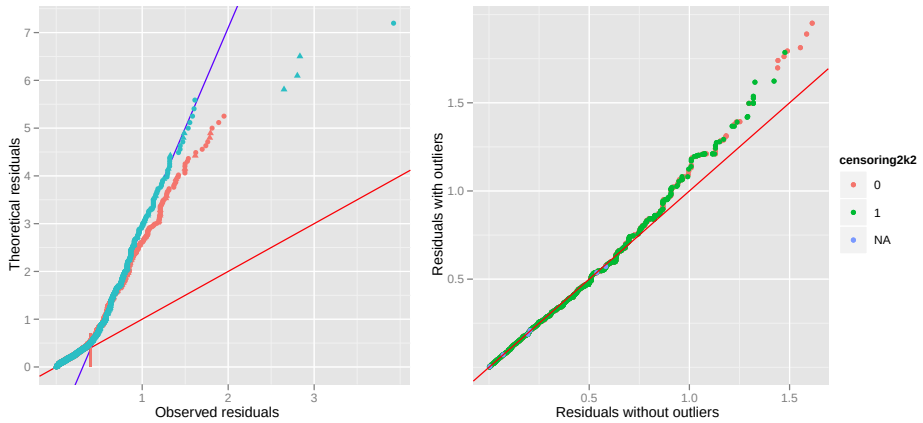
Figure 5.6: Cox-Snell residual contrasts between Kim, Cox, Cowling and Rogers models.

5.3 Residuals without outliers

From the residual plots observed previously, Kim and Cox models indicate six possible outliers. Since it is still under discussion whether or not the residuals with greatest values signaled by Cowling and Rogers models are to be considered as

outliers themselves, we restrict this section to observe the difference of the models outcomes if the six Kim and Cox's models outliers were taken out from the study.

We start by comparing the outcomes under Kim's model for the full data set and the new data set without the six outliers. In Table 5.3 we observe, for this model, that there is little variation for the estimates of EEG and neurological disorders. It is the logarithm of the number of seizures that shows a more substantial increase in the estimated coefficient, and such an effect becomes clearer in Figure 5.7. The increase in this estimate was to be expected, since we are modifying the data set precisely by taking out patients with very high number of seizures, relative to the rest of the population. Indeed, Kim's model signals patients with large numbers of seizures as outliers more frequently than any of the other models. For the likelihood ratio test, where we consider the current likelihood over the saturated-model likelihood, we obtained a ratio of 77.8 for the full data set, and a ratio of 96.3 for the reduced data set. Since these quantities are based in different population sizes, we could have that the likelihood ratio under the reduced data is greater simply because it contains less data. For this reason we normalize both quantities by dividing them by the respective number of patients they are considering. This leaves us with the normalized likelihood ratio statistics $77.8/1334 = 0.058$ and $96.3/1328 = 0.072$ respectively. This shows that the model under the reduced data set is more plausible than the model under the full data set. This type of likelihood ratio statistic normalization will be carried out for the remaining three models as well.



(a) Q-Q plots (b) Contrast of the model

Figure 5.7: Kim model's Cox-Snell residual contrasts between the data sets with and without outliers.

Consider Figure 5.7(a) where we observe the Q-Q plots for the Cox-Snell residuals contrasted to an $Exp(1)$ distribution. The residuals for the full data set are shown in blue, whilst the new residuals are shown in red. Censored data and observed data are represented by triangles and circles respectively. Although previously we had observed that the Cox-Snell residuals r_C seemed to follow the form $r_C = -1.3 + 4.2 \times Y$, where $Y \sim Exp(1)$, the new residuals do not fit the same regression line. Indeed the slope seems to be smaller, therefore fitting the assumptions slightly better. In Figure 5.7(b) the Cox-Snell residuals for the same model and different data sets are contrasted against each other. Here censored and observed data are plotted in green and red colors respectively. Although the majority of the early-seizure residuals tend to be the same, as the time to first seizure increases the value of the newly-fitted residuals also increase with respect to their former values.

	Full Data			Reduced Data			Ratio
	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	p-value	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	p-value	$\hat{\beta}_1/\hat{\beta}_2$
Ndl	0.40	0.12	$1.3e^{-3}$	0.42	0.13	$7.8e^{-4}$	0.95
EEG	0.31	0.08	$7.9e^{-5}$	0.34	0.08	$1.8e^{-5}$	0.92
log(No.seiz)	0.38	0.05	$6.6e^{-15}$	0.52	0.06	$1e^{-15}$	0.73
Likelihood ratio/n			0.058			0.072	

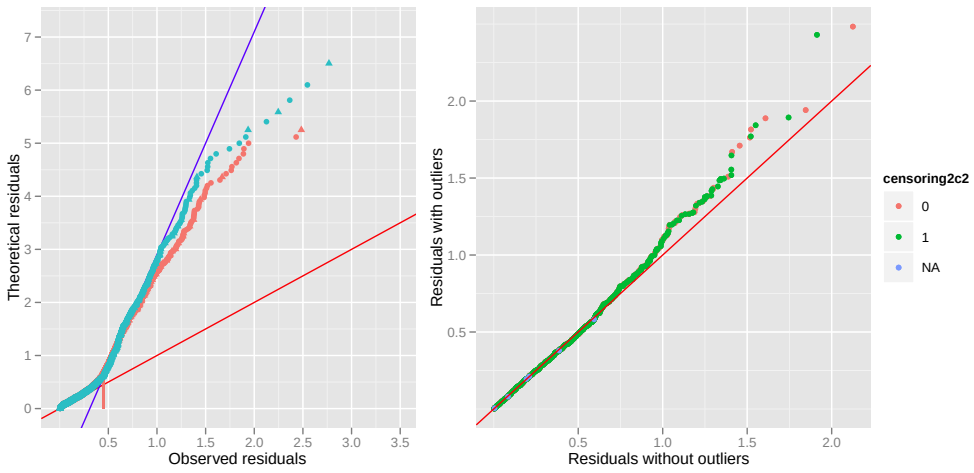
Table 5.3: Kim's model fitted coefficients for the data sets including and excluding the outliers, respectively.

A similar phenomena is observed for the Cox model, where Figure 5.8 represents censored and observed data as mentioned before. In Figure 5.8b) the residuals under the new data set are smaller than the residuals for the full data, and in 5.8a) these new residuals present a slightly better fit under the assumptions. Observe that the new residuals no longer approximate the previously fit line passing through 0 and slope 2.1.

In Table 5.4, we observe that the estimated regression coefficients for the two data sets vary more noticeably for the EEG and Treatment covariates and their interactions with the Tonic-Clonic seizure indicator. This could be a result from excluding six patients of whom one had partial seizures, and the remaining presented 2nd. Tonic-Clonic seizures. As before, the likelihood ratio statistics are normalized by the number of patients under consideration for each case, and these are found to be $108/1334 = 0.08$ and $129/1328 = 0.09$ for the full and reduced data sets respectively. Observe that although there are some noticeable changes in the estimated coefficients, the likelihood ratio statistics show that the model under the reduced data set does not present a large improvement over fitting the model to the full data set.

	Full Data			Reduced Data			Ratio
	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	p-value	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	p-value	$\hat{\beta}_1/\hat{\beta}_2$
Trt	0.31	0.28	0.26	0.22	0.28	0.42	1.39
T-C	-0.25	0.27	0.36	-0.24	0.26	0.36	1.04
2nd.T-C	-0.38	0.28	0.17	-0.37	0.27	0.18	1.03
EEG	0.08	0.28	0.79	0.20	0.28	0.46	0.37
Age	-0.005	0.002	0.66	-0.001	0.002	0.48	0.62
log(No.seiz)	0.38	0.05	<0.005	0.50	0.06	<0.005	0.75
T-C×EEG	0.36	0.29	0.21	0.25	0.29	0.38	1.43
2nd.T-C×EEG	0.83	0.30	0.01	0.76	0.30	0.01	1.09
Trt×EEG	-0.55	0.16	<0.005	-0.59	0.16	<0.005	0.92
Trt×T-C	-0.38	0.28	0.18	-0.27	0.28	0.33	1.39
Trt×2nd.T-C	-0.45	0.29	0.12	-0.37	0.29	0.21	1.22
Likelihood ratio/n			0.08			0.09	

Table 5.4: Cox's model fitted coefficients for the data sets including and excluding the outliers, respectively.



(a) Q-Q plots (b) Contrast of the model

Figure 5.8: Cox model's Cox-Snell residual contrasts between the data sets with and without outliers.

Although for the Cox and Kim models the residual analysis for the full and reduced data sets showed little improvement, in the following analyses for the Cowling and Rogers models we observe even smaller changes. In the following Table 5.5 we show the covariate coefficient estimates for Cowling's model, for both data sets

under consideration. Observe that the heterogeneity measure α does not change significally, as well as the type of seizure covariates in the pre-randomization rate. For the rest of the covariate coefficients, however, they usually are doubled for the full data set with respect to the reduced data set. In particular, the Tonic-Clonic seizure factor plays a less significant role in the post-randomization seizure rate, as it decreases by ten times relative to the same coefficient for the full data set, although its interactions with other factors are not decreased in the same magnitude. It would seem that the omission of the six patients with partial and 2nd. Tonic-Clonic seizures leads to a substantial change in the model, in which having a Tonic-Clonic seizure leads to a much greater decrease in the seizure rate than before. For this post-randomization rate, the coefficient estimates for both Tonic-Clonic seizure factor and its interaction with the EEG outcome are diminished to the point of becoming negligible. Indeed, their standard errors are greater or equal to their estimated values, meaning that under the reduced model, we could consider to simplify the model by taking these two factors out of the analysis.

Although the covariate coefficients tend to change more noticeably in this model than in Kim or Cox's models, by observing the resulting standarized log-likelihoods, the model under the reduced data set fits the data almost in the same way as for the full data set, if only slightly better. This, in turn, can be corroborated when observing the contrast plot in 5.9b), where the residuals appear to be smaller for the reduced data set, in a more consistent form than for the previous models. In Figure 5.9a) we observe that the new residuals fit the assumptions somewhat better, although they no longer fit the line with intercept 0 and slope 2.1. They still, however, deviate from the expected residuals from the beginning.

When considering Rogers model, in Table 5.6 the coefficient estimate changes behave in a very similar way for the pre-randomization seizure rate as the changes observed in Cowling's model. The intercept and types of epilepsy do not present a significant change, although the rest of the coefficients tend to reduce by half their

		Full Data		Reduced Data		Ratio
		$Coef_1$	SDs	$Coef_2$	SDs	$Coef_1/Coef_2$
α		1.99	0.11	2.08	0.12	0.95
β_1	Intercept	-4.54	0.15	-4.54	0.15	1.00
	T-C	-0.69	0.16	-0.69	0.16	0.99
	2nd T-C	-0.40	0.17	-0.40	0.16	0.97
	EEG	0.57	0.19	0.30	0.19	1.89
	Age	-0.003	0.001	-0.002	0.001	1.58
	T-C*EEG	-0.53	0.20	-0.26	0.20	2.07
	2nd T-C*EEG	-0.33	0.21	-0.15	0.21	2.17
β_2	Intercept	-2.61	0.33	-2.26	0.31	1.15
	Trt	0.99	0.33	0.35	0.33	2.80
	T-C	0.41	0.34	0.04	0.32	10.13
	2nd T-C	0.15	0.35	-0.22	0.34	-0.70
	EEG	-0.47	0.35	0.19	0.34	-2.47
	Trt*T-C	-1.22	0.33	-0.54	0.33	2.26
	Trt*2nd T-C	-1.34	0.35	-0.67	0.34	2.00
	Trt*EEG	-0.64	0.18	-0.73	0.18	0.87
	T-C*EEG	0.97	0.36	0.35	0.35	2.77
	2nd T-C*EEG	1.57	0.37	1.16	0.37	1.35
	-LogLikelihood/n		5.89		5.82	

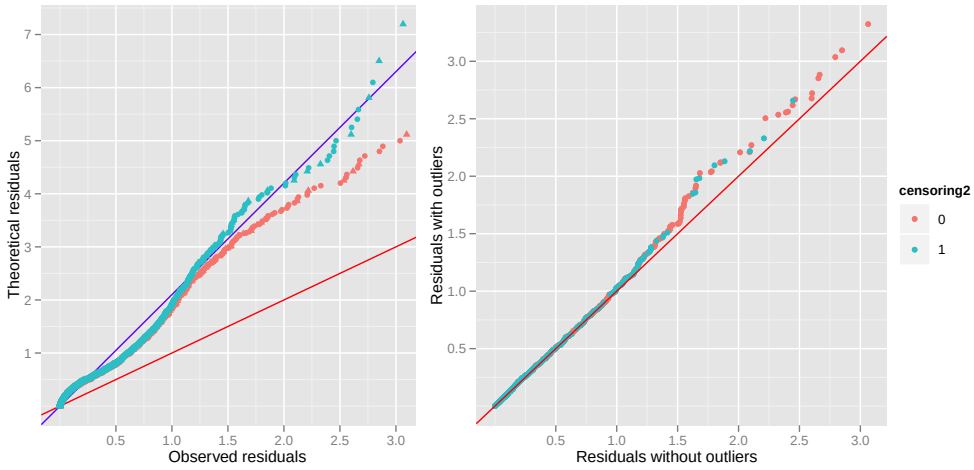
Table 5.5: Cowling model's fitted values for the full data set versus the data set without outliers.

value for the reduced data set. For the post-randomization seizure rate, however, the coefficient estimates remain generally unchanged, except for the treatment covariate and its interactions. In these cases the coefficients tend to decrease by more than one third, but notice that the standard errors are of such magnitudes, that these factors would lose influence in the seizure rate. The cure fraction, on the other hand, shows generally increased coefficients for the new data set, save for the EEG outcome covariate. For the EEG the coefficient decreases more than twice. This decrease and its standard error size dims the EEG factor negligible to the model, in contrast to its interaction with the treatment effect, in which case this interaction becomes more influential to the cure rate.

		Full Data		Reduced Data		Ratio
		$Coef_1$	SDs	$Coef_2$	SDs	$Coef_1/Coef_2$
α		2.08	0.11	2.24	0.13	0.927
β_1	Intercept	-4.53	0.15	-4.54	0.15	1.00
	T-C	-0.68	0.16	-0.69	0.16	0.99
	2nd T-C	-0.38	0.16	-0.40	0.16	0.95
	EEG	0.56	0.18	0.29	0.18	1.91
	Age	-0.004	0.001	-0.003	0.001	1.54
	T-C*EEG	-0.54	0.20	-0.26	0.20	2.03
	2nd T-C*EEG	-0.34	0.21	-0.15	0.21	2.32
β_2	Intercept	-1.06	0.38	-0.95	0.36	1.12
	Trt	0.39	0.39	0.17	0.38	2.26
	T-C	0.57	0.39	0.51	0.38	1.12
	2nd T-C	0.55	0.40	0.50	0.39	1.10
	EEG	0.59	0.40	0.52	0.40	1.14
	Trt*T-C	-0.60	0.41	-0.48	0.40	1.23
	Trt*2nd T-C	-0.89	0.41	-0.55	0.40	1.62
	Trt*EEG	-0.78	0.26	-0.45	0.25	1.73
	T-C*EEG	-0.56	0.42	-0.60	0.42	0.94
	2nd T-C*EEG	-0.44	0.43	-0.31	0.43	1.41
κ_1	Intercept	0.88	0.50	0.80	0.48	1.09
	Trt	0.10	0.50	0.21	0.50	0.47
	T-C	-0.82	0.51	-0.78	0.49	1.05
	2nd T-C	-1.06	0.52	-0.99	0.51	1.08
	EEG	-0.30	0.52	-0.12	0.52	2.55
	Trt*T-C	-0.18	0.50	-0.22	0.50	0.81
	Trt*2nd T-C	-0.11	0.53	-0.31	0.52	0.35
	Trt*EEG	-0.40	0.30	-0.64	0.28	0.62
	T-C*EEG	0.96	0.54	0.86	0.53	1.12
	2nd T-C*EEG	1.78	0.57	1.62	0.55	1.10
	-LogLikelihood/n		5.72		5.66	

Table 5.6: Rogers model's fitted values for the full data set versus the data set without outliers.

When observing the standarized log-likelihoods, we observe that there is little change in the likelihood fits provided by the model under the two data sets. This assessment is reinforced by the plots in Figure 5.10, where both the QQ-plot and the contrasts plot show a negligible change between residual estimates. A change so small in the expected residuals shows that from the four models under consideration,



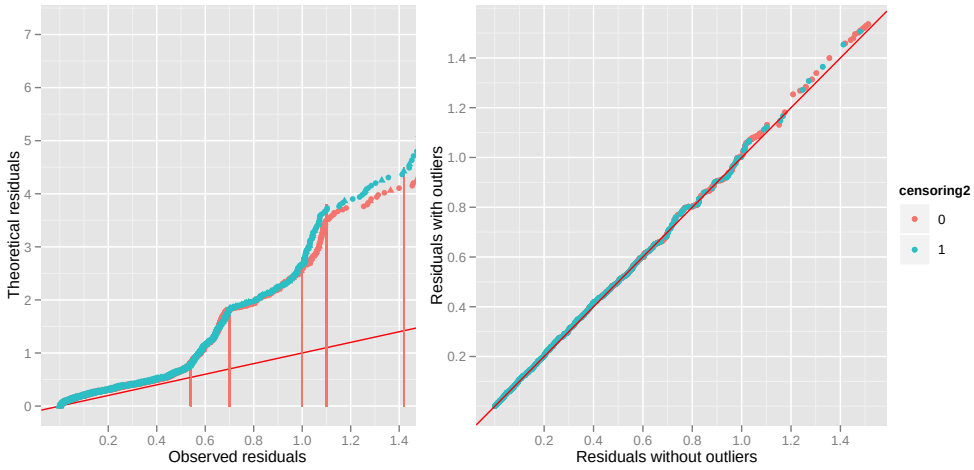
(a) Q-Q plots (b) Contrast of the model

Figure 5.9: Cowling model's Cox-Snell residual contrasts between the data sets with and without outliers.

Rogers model is the most robust to the data set's extreme values for covariates such as number of seizures and seizure rate. We hold in consideration that, although the omittance of the six outliers naturally benefits Kim and Cox models more clearly than for Cowling and Rogers models, the residual analysis still shows a significant departure from assumptions for all models. We must have under consideration, however, that literature discusses the interpretation of the departure of residuals from the expected values, and a more thorough study of residual analysis is required. For this purpose, in the following section we produce survival data simulations and fit them under two of these models.

5.4 Residual study for negative binomial residuals

In the previous subsections we have presented and showed the residual studies for the first post-randomization seizure times Y_1 under the assumption of four different models. Let us now consider the pre-randomization seizure count X_i and remember that, under both the joint model and the joint model with cure fraction, we consider



(a) Q-Q plots (b) Contrast of the model

Figure 5.10: Rogers model's Cox-Snell residual contrasts between the data sets with and without outliers.

the seizure count to be distributed as a negative binomial. We are interested in observing, on one hand, how do the residuals under this distribution perform, for instance, in the case of the Cox-Snell residuals against an exponential distribution with mean one. Secondly, we wish to compare the residual allocation for the seizure count, in contrast to the residual allocation for the seizure time for each individual, and observe if they project consistent behaviours.

Let us consider the joint model for the epilepsy data. In the left-side of Figure 5.11 we show the plot of the Cox-Snell residuals for the seizure counts, along the horizontal axis, against the respective residuals for the seizure times, along the vertical axis. The colours denote the type of epilepsy, where red is Tonic-Clonic only, green is 2nd generalized and blue are partial seizures. The circles denote an observed seizure time, whilst the triangles denote censored seizure times. In this plot we observe that the majority of the individuals under observation produce small residuals under the negative binomial distribution. The residuals for the same individuals under the Lomax distribution, however, appear to project a more uniform span for the residuals.

The difference in the negative binomial survival functions between the joint and the joint model with cure, lies in the difference between their estimated pre-randomization parameters α and λ_i, since the seizure count considered is recorded pre-randomization. For this reason, the corresponding residual analysis between the joint and joint model with cure are similar. In the left-side of Figure 5.11 we present the Exponential Q-Q plot for the ordered Cox-Snell residuals under the joint model, whilst the right-side of the Figure shows the corresponding residuals under the joint model with a cure fraction. As we had observed before, it is apparent that the distribution of the seizure counts diverges from a negative binomial distribution. There appears to be no specific seizure type pattern along the curve, nor does it appear to present a censoring particular behaviour. The blue line represents a fitted line and in these cases we consider the line to have intercept 0 and slope 3.7 for both models. Observe that 97.45% for the joint model and 97.37% for the model with a cure fraction of the Cox-Snell residuals follow the pattern of this fitted line, and the remaining proportion corresponding to the greatest residuals in value digress under the fitted line. In these cases we observe that the four highest valued residuals do not correspond to the same individuals signaled as possible outliers by the same models under the Lomax distribution.

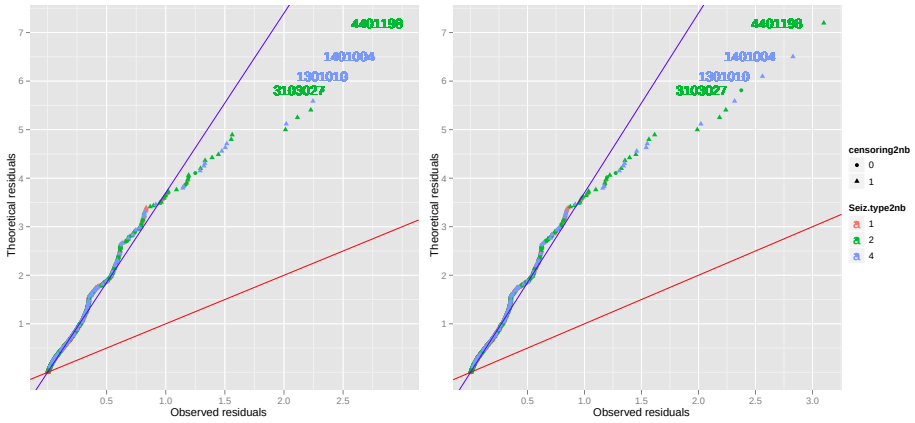
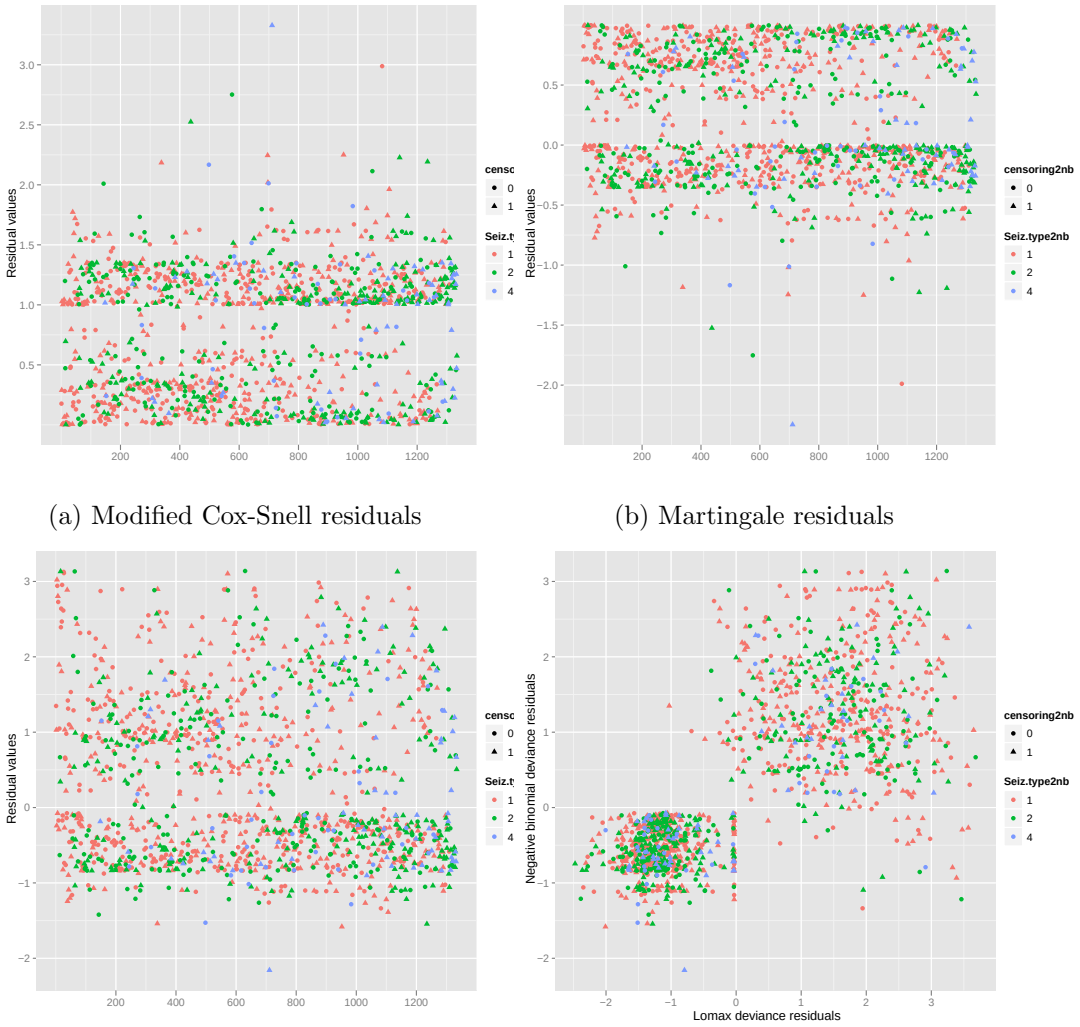


Figure 5.11: Q-Q plots of residuals under the negative binomial distribution for the joint and joint (left) with cure fraction (right) models.

In Figure 5.12 and Figure 5.13 we present the Modified Cox-Snell, Martingale and Deviance residuals for the joint and joint with cure fraction models respectively. For both models, when observing the Modified Cox-Snell (top-left) and Martingale (top-right) residuals, we observe no particular pattern in their layout, and although there are a few residuals that digress from the center, the majority of the residuals are clustered about zero and unity. In the Deviance residuals (bottom-left) there is no particular pattern in the index plot, and as seen before for the corresponding Lomax residuals, they do not appear to be centered at zero, but rather, these index plots show a heavier right tail. The last plot represents the contrast of the deviance residuals under the Lomax and negative binomial distributions for the same model, and are shown in the bottom-right side of the respective Figures. Observe that the left tail of the deviance residuals tend to be similarly categorized for the count and the seizure times. The right tail of these residuals also tend to categorize the residuals in a consistent way, except for a lower number of residuals corresponding majorally to patients with Tonic-Clonic seizures only, or second generalized seizures, whilst the individuals with partial seizures only tend to be very consistently categorized for time and seizure counts.



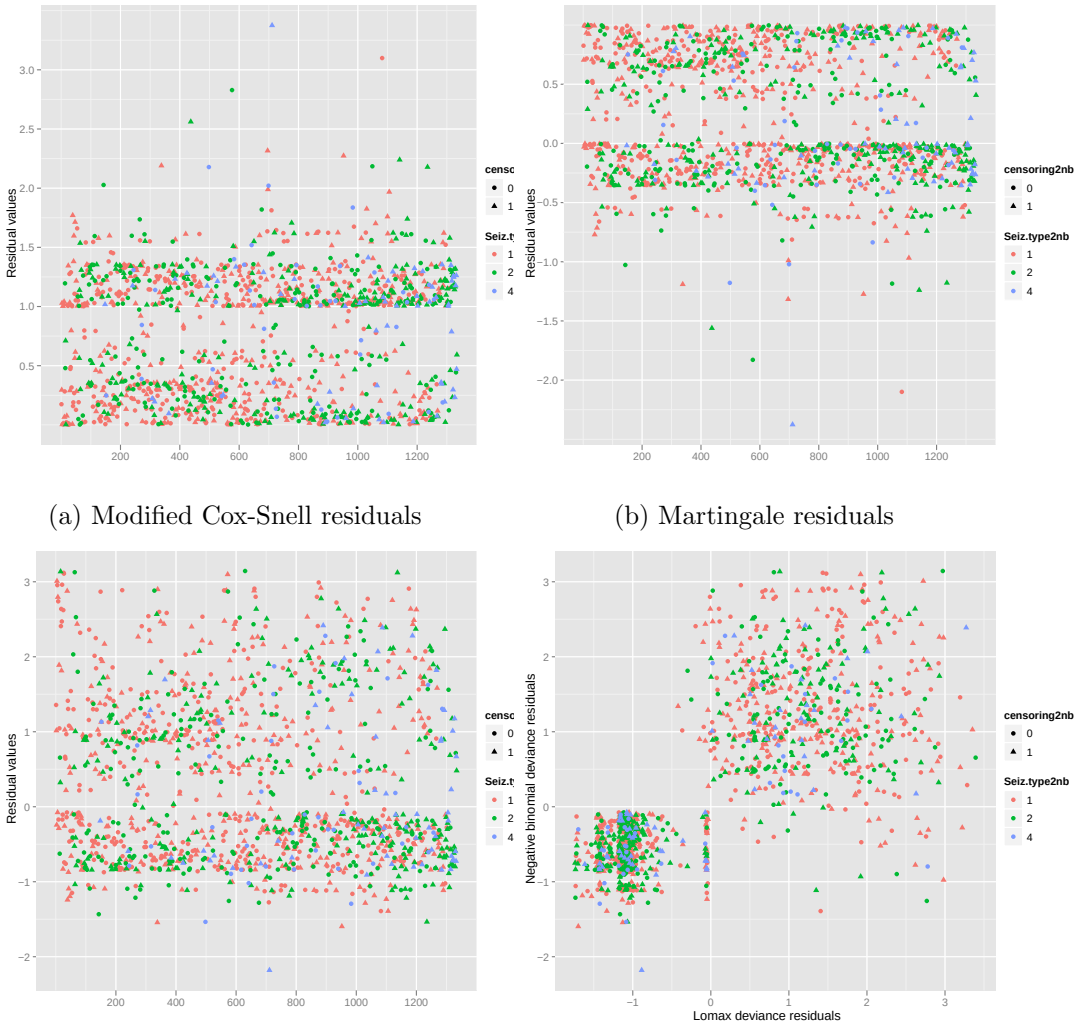
(a) Modified Cox-Snell residuals

(b) Martingale residuals

(c) Deviance residuals

(d) Contrast of the Deviance residuals for time versus seizure count

Figure 5.12: Residual plots for the seizure counts under the joint model.



(a) Modified Cox-Snell residuals

(b) Martingale residuals

(c) Deviance residuals

(d) Contrast of the Deviance residuals for time versus seizure count

Figure 5.13: Residual plots for the seizure counts under the joint model with a cure fraction.

Let us remember that the only semi-continuous covariate we can consider is the age of the patient at randomization. In the previous figures we have shown the index plots, but the residual layout is in part aimed at finding possible covariate influences that have not been properly addressed by a model. For this reason in Figure 5.14 we plot the Deviance residuals of the joint (left side) and the joint model with cure fraction (right side). In these plots we observe that there is a

noticeably higher concentration of the residuals for younger patients, particularly under 30 years of age. Let us remember that the age of the patients at randomization has a mean of 30 years, ranging from less than a year old to 62 years of age, and that the ages have been rescaled so that they are centered at 30 years. From these plots we have an indication that the population of patients under 30 years of age could be studied and compared to the results from the population who's age is greater than 30. We address this topic in Section 5.8.

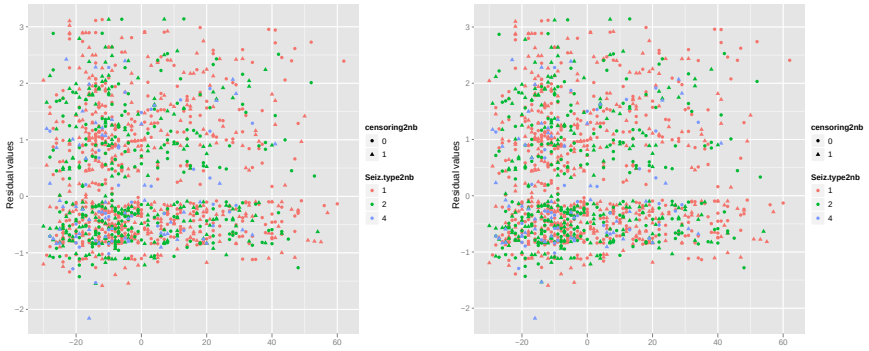


Figure 5.14: Deviance residuals for seizure counts under the joint (left) and joint with cure fraction (right) models.

5.5 Residual study of simulated data

There is little guidance in the literature on the use of residuals to examine closely related models. From the residual analysis shown previously we observe that the Q-Q plots for the Cox-Snell residuals indicate a certain proportion of well fitted values, but the patients with later seizures seem to deviate from the model's assumptions in each case. In an attempt to understand the interpretation of the obtained residuals, we proceed to generate a number of sets of simulated data and contrast the fits shown under Cox's proportional hazards model, and under Cowling's model.

In order to model populations with small and large heterogeneity, we produced two populations of simulated data considering dispersion parameters $\alpha_1 = 0.5$

and $\alpha_2 = 4$. For each population we consider 15 sub-populations, each one with a sample size of 1000 individuals. The number of seizures X_i and first seizure time T_i for each individual i are modeled from a Poisson-gamma mixture distribution. The conditional seizure number $X_i|\nu_i$ is generated from a Poisson distribution with parameter $\nu_i\lambda_i u_i$, where $\nu_i \sim Gamma(\alpha, \alpha)$ and u_i is the period of pre-randomization observation. Here we set all pre-randomization observation periods to be 182 days, according to the minimum period considered for the MESS study. For the pre-randomization seizure rate λ we considered in each sample 500 individuals having a rate $\lambda_1 = 2$ and the remaining 500 patients having a rate $\lambda_2 = 6$. The purpose of this mixing is to emulate the different rates that could be observed in a population of epilepsy patients that have different types of epilepsy syndromes, such as Tonic-Clonic and Partial seizures. We must consider, in addition, that a patient is not diagnosed with epilepsy unless he has had at least one seizure, and hence the number of seizures were zero-truncated, a consideration that we keep in mind for later generalizations of the model.

The conditional times to first seizures $T_i|\nu_i$ are generated from an exponential distribution with parameter $\lambda_i\psi_i\nu_i$, where we consider two post-randomization rate modifiers $\psi_1 = 0.35$ and $\psi_2 = 0.65$. For each of the 15 sub-populations we then generate seizure times from combining the two values of λ and the two values of ψ. Finally, for these simulations we produce a censoring mechanism of the times to first seizures. Since in real survival data we would expect the early-occurrences to be observed, and have censored observations after a certain threshold of the period, we select the 40% quantile $t_{(40)}$ of each sample $t_1, t_2, ..., t_{1000}$, and for all $t_i \geq t_{(40)}$, we produce the censored seizure time t_i^* from the same distribution the observed seizure time t_i was simulated. Once we obtain these simulated censored times, we create the final sample of seizure times y_i by selecting $y_i = min\{t_i, t_i^*\}$.

We now propose to contrast the residual fits between four proportional hazards models and Cowling's model. Let us denote by Z and W the indicator vectors

for the values of λ and ψ respectively. We are interested in contrasting residual analyses provided by considering all the covariate combination possibilities. We will fit the simulated data to the following models:

$$
\begin{aligned}
Model\ 1 \quad &: \quad h(y) = h_0(y)\exp(b_1 z + b_2 w), \\
Model\ 2 \quad &: \quad h_z(y) = h_{z0}(y)\exp(b_2 w),\ \text{stratifying by } z, \\
Model\ 3 \quad &: \quad h(y) = h_0(y)\exp(b_1 z + b_2 w + b_3 log(x)),\ \text{with } x \text{ the number of seizures}, \\
Model\ 4 \quad &: \quad h_z(y) = h_{z0}(y)\exp(b_2 w + b_3 log(x)),\ \text{stratifying by } z,
\end{aligned}
$$

where b_1, b_2 and b_3 are the regression coefficients. Finally we consider Model 5 to be Cowling's model, which proposes that $X|\nu \sim Poi(\lambda\nu u)$ and $Y|\nu \sim Exp(\lambda\psi\nu)$, with $\lambda = \exp(b_1 Z)$ and $\psi = \exp(b_1 Z + b_2 W)$.

Let us remember that for a Cox proportional hazards framework, we assume that the general form of the hazard function is of the form

$$h(t) = h_0(t)\exp(b_1 z_1 + b_2 z_2 + ... + b_p z_p),$$

when considering p predictor variables $z_1, z_2, ..., z_p$, and where $b_1, b_2, ..., b_p$ are the regression coefficients. Hence we have that the logarithm of the hazard ratio will consequently leave us with the linear combination expression

$$\log\left(\frac{h(t)}{h_0(t)}\right) = b_1 z_1 + b_2 z_2 + ... + b_p z_p,$$

and when a predictor z_i is dichotomous, its corresponding regression coefficient estimate $\exp(\hat{b_i})$ represents the increase of the risk event for the individuals with $z_i = 0$ over the risk for individuals with $z_i = 1$. In our case, observe that in all four Cox models, both z and w predictor variables are dichotomous. A direct comparison to the real values of the predictor variables can be attained by recalling that the times to first seizures were modeled from an exponential function, which means that

the hazard function is a constant; more precisely, is the seizure rate. A hazard ratio for the pre-randomization rate will be, in this case, be given by

$$\frac{h_{z_1=0}(y)}{h_{z_1=1}(y)} = \frac{h_{\lambda_1}(y)}{h_{\lambda_2}(y)} = \frac{\lambda_1}{\lambda_2} = \frac{2}{6} = 0.33.$$

In a similar way, for the post-randomization seizure rate, we have that the corresponding hazards ratio is of the form

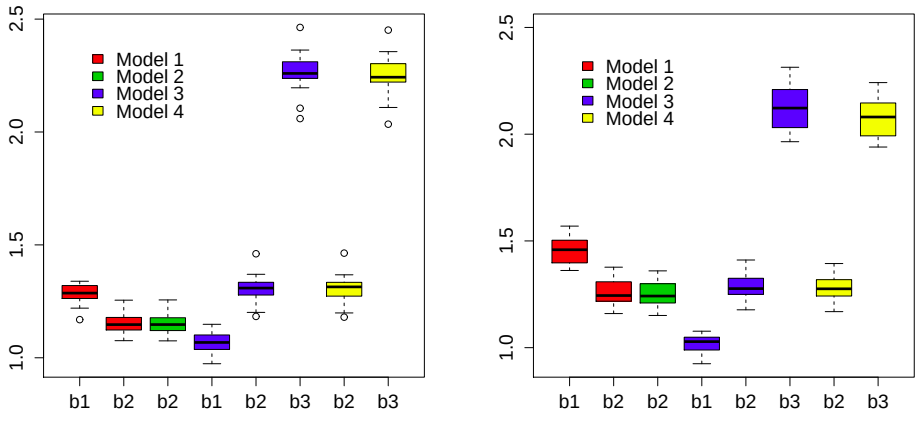
$$\frac{h_{w_1=0}(y)}{h_{w_1=1}(y)} = \frac{h_{\psi_1}(y)}{h_{\psi_2}(y)} = \frac{\psi_1}{\psi_2} = \frac{0.35}{0.65} = 0.54.$$

In Cox proportional hazards models, the interpretation of the regression coefficients is such that, if for a predictor z_i the corresponding estimated coefficient $\hat{b_i}$ is close to one, then the predictor in question does not affect the survival. If such coefficient, however, is less than one, the predictor improves the survival, and conversely it increases the risk when the coefficient is greater than one.

Each of the five models was fitted to the two populations, leaving us with estimated coefficient values for 15 sub-populations considering α_1 and 15 populations considering α_2 to be the underlying heterogeneity rates. We obtain the regression coefficient estimates $\hat{b_{1j}}$, $\hat{b_{2j}}$ and $\hat{b_{3j}}$ with $j = 1, ..., 15$, and which correspond to the predictors z, w and $log(x)$ respectively, where x is the number of seizures of the patient. For simplicity, we represented the estimated regression coefficients in a graphical manner, omitting their corresponding standard errors. The 15 sets of regression coefficients for the first four models are shown in 5.15, where each boxplot corresponds to the estimated hazard ratios ($\exp(\hat{b_{kj}})$, $k = 1, 2, 3$ and $j = 1, ..., 15$) under a specific model. The boxplots are distinguished by color, where Models 1 to 4 are represented by red, green, blue and yellow respectively.

Observe in Figure 5.15 that, although the scales are slightly different for the boxplots under the two populations, they share the same trends for all four models. Considering that the original hazard ratios for z and w are 0.33 and 0.54 respectively,

we observe that the estimated hazard ratio boxplots in general do not cover such values. In particular, Model 1 shows that the pre-randomization seizure rate has a greater impact on the risk of future incidences than the post-randomization modifier, or in other terms, than the treatment effect itself. Observe that for both Figure 5.15a) and Figure 5.15b), Model 2 does not consider the interaction $z \times w$ in the regression, and it produces undistinguishable hazard ratios for the predictor w as Model 1, suggesting that, independently of the value of λ, the patients with a $\psi_1 = 0.35$ change in the seizure rate will experience from 1.1 to 1.3 times a higher risk than those individuals with a $\psi_2 = 0.65$ change in the rate. A similar behaviour is observed between Models 3 and 4, where the lack stratification with respect to z in Model 4 produces very similar hazard ratios for the remaining factors w and $\log(x)$. It is worth noticing that the hazard ratios for z for both populations attain or are very close to a unit value, which suggests that the factor in question has little or no effect on the incidence risk. Once we consider the number of seizures per patient, their underlying pre-randomization seizure rate is deemed negligible. In fact, the logarithm of the number of seizures has a much greater impact on the seizure recurrence. Indeed, the predictor $\log(x)$ has an impact from 2 to 2.5 for Model 3 and from 2 to 2.4 for Model 4 in the hazard ratio. The main difference between the populations lies in that the population with a higher heterogeneity presents a slightly lower dispersion with higher median values for this factor, and the population with a lower heterogeneity does not show any extreme values. From the plots we observe that the presence of a larger heterogeneity tends to produce more concentrated estimates and deems some observations as extreme values.



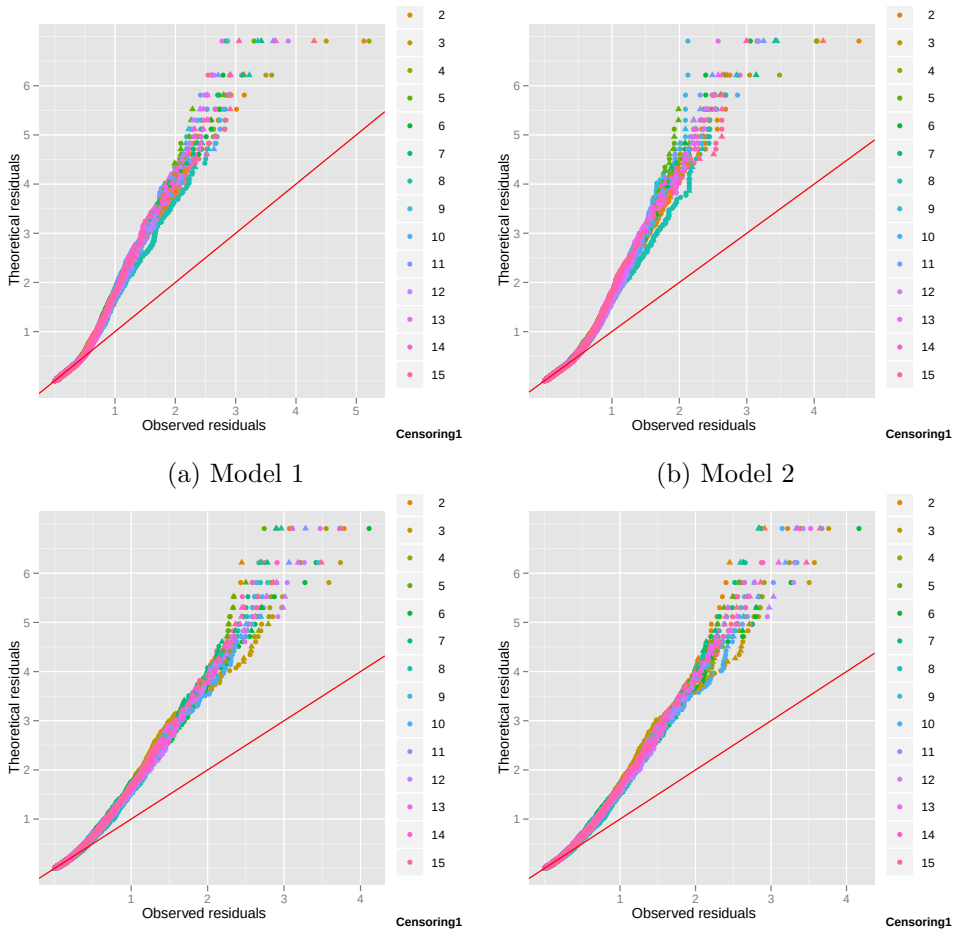
(a) For $\alpha_1 = 0.5$ (b) For $\alpha_2 = 4$

Figure 5.15: Box plots for the 15 sets of estimated hazard ratios, under the four Cox models proposed.

Although in Figure 5.15 the plots seem to indicate very similar regression fits between the two populations, we are interested in contrasting their respective fits in terms of their residual analyses. In Figure 5.16 and Figure 5.17 we present the resulting Cox-Snell residual Q-Q plots for the Cox proportional hazard models, corresponding to the underlying value of α_1 and α_2 respectively. Each of the 15 sub-populations is indicated by a different colour, where circles and triangles correspond to observed and censored seizure times respectively.

From the previous boxplots we observed that the estimated hazard ratios for the first four models do not resemble the true hazard ratios from which the data set was produced. This same lack of fit is observed in the residual analyses shown in Figure 5.16. Although the early seizure time residuals are well fitted, the rest of the residuals quickly diverge from their expected values in a linear form. Indeed, the observed residual values tend to be half of the expected values, underestimating the true expected seizure times. The large heterogeneity seems to discriminate better which regression coefficients the Cox model will estimate, and this in turn produces

very similar residual sub-populations within the models. In fact, we only observe some dispersion between the sub-population residuals for approximately the last 5% of late occurrences. Observe that the observed residuals tend to lie within the same range of values for the four models, and there appears to be no substantial difference in the behaviour of residual sub-populations between models, which suggests that the residual analyses deem all four models to be inadequate in very similar ways for this population.



(a) Model 1

(b) Model 2

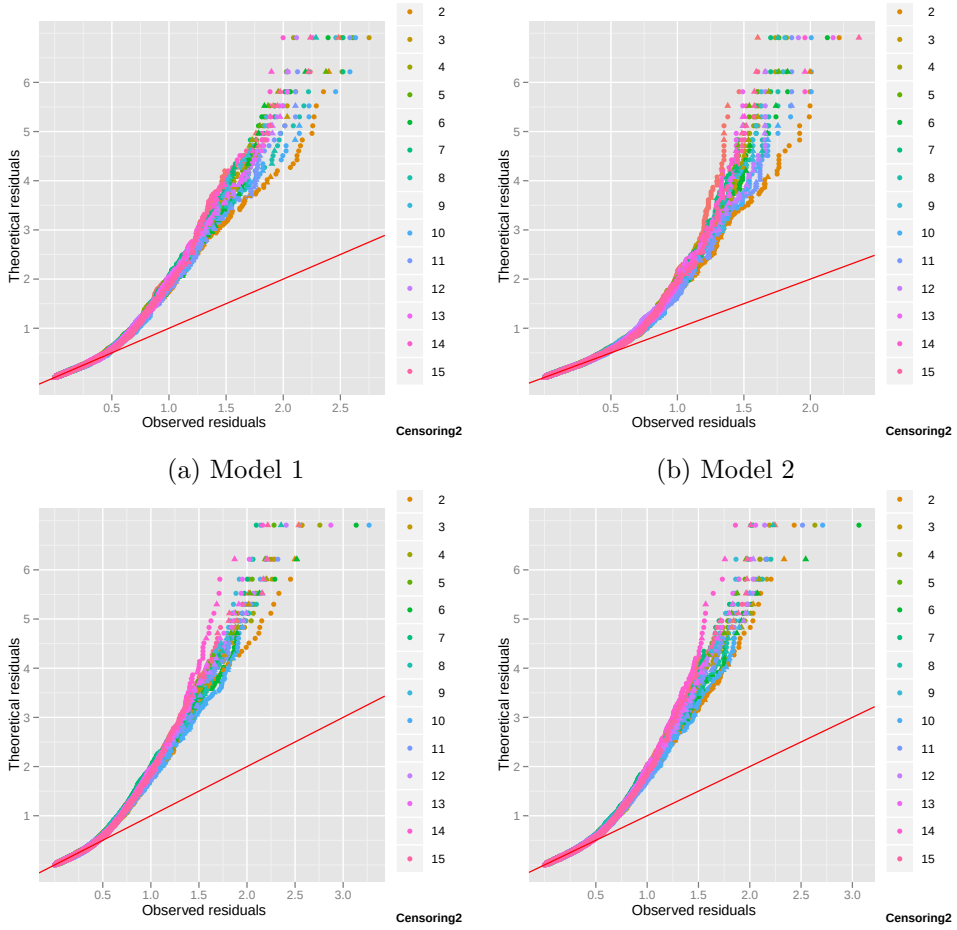
(c) Model 3

(d) Model 4

Figure 5.16: Q-Q plots of the Cox-Snell residuals for proportional hazards models. $\alpha = 0.5$ is the underlying heterogeneity rate.

When considering a smaller heterogeneity in the population, the residual fits

confirm once again the slightly greater dispersion of estimates that was observed in the boxplot figures. In the same way we observed for the population with underlying α_1, the population with α_2 shown in Figure 5.17 produces well fitted residuals under the Cox models, but only for the very early seizure incidences. Later on the sub-population residuals present a greater sparcity than that observed for largely heterogeneous populations. Notice that for the population under α_2 the observed residual values range within a smaller interval than those observed in Figure 5.16, which shows a steeper departure from the expected residual values. In this case the sub-population residuals underestimate the expected values in a less linear form, tending to an exponential shape as it is more clearly observed for Models 2 and 4, which happen to be the models stratified by z. This provides an insight that the inclusion of the z factor in the Cox model provides a slightly better fit for this kind of population, as opposed to omitting it. However, as it was noticed before, the Cox proportional hazards model provides a poor fit for this kind of data. What is interesting is that the joint models do show, for the two populations considered, a very similar behaviour as the one observed for the Kim and Cox models considered for the MESS data set. Let us remember that the Q-Q plots observed in Figure 5.1 fitted well the early seizure times, and then underestimated the expected values in a linear form. This in principle might signal that Cox proportional hazards model fits the data in a very similar way that it fits for the MESS data set, and that hence the simulated and MESS data sets are very similar in nature, or perhaps that there are other underlying features of the data that the model is not accounting for.



(a) Model 1

(b) Model 2

(c) Model 3

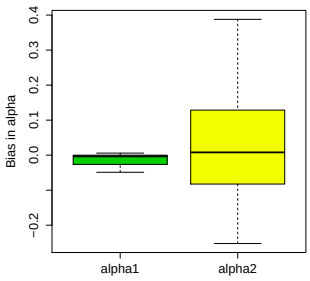
(d) Model 4

Figure 5.17: Q-Q plots of the Cox-Snell residuals for proportional hazards models. $\alpha = 4$ is the underlying heterogeneity rate.

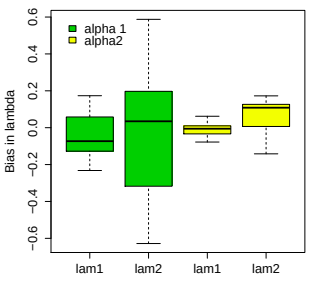
In an attempt to investigate further the nature of the residual analysis, we proceed to fit another model to this simulated data. We now fit Cowling's model, and since the simulations were designed much in the same framework that Cowling's model assumes, we expect the estimated likelihood coefficients to be very similar to the original coefficients used. Once again we fit the model for both populations with underlying values of α_1 and α_2. For each sub-population within the heterogeneity group we obtain numerically the estimates for the parameters α, λ and ψ, and we display them in a graphical form by means of plotting the boxplots of their respective

biases. In Figure 5.18a) we show the bias corresponding to $\alpha_1 - \hat{\alpha_1}$ in green, and the bias $\alpha_2 - \hat{\alpha_2}$ in yellow. Although the values for both cases are clearly centered at zero, the presence of a higher heterogeneity appears to produce a much better identification of the true value of α for this model, as shown by the difference in dispersion for the boxplots in the figure. This property takes the converse order, however, when measuring the bias between the true and the estimated values of λ_1 and λ_2 under the two underlying values of α. Observe in Figure 5.18b) the corresponding biases of $\lambda_i|\alpha_j - \hat{\lambda_i}|\alpha_j$ for $i, j = 1, 2$. It is apparent that, although the higher the level of heterogeneity makes it easier for the model to estimate α, it also causes that with a large heterogeneity there are numerous types of pre-randomization rates in the population, making their estimation increasingly variable, although still taking values in the vicinity of the true value of the parameter. Observe that even when the dispersions of the boxplots vary, the medians are always close to zero. For the last parameter ψ, the biases of the form $\psi_k|\alpha_j, \lambda_i - \hat{\psi_k}|\alpha_j, \lambda_i$ for $k = 1, 2$, are shown in Figure 5.18c). Notice that the biases corresponding to the underlying parameter α_2 tend to deviate slightly more from zero than those corresponding to α_1, the scale is sufficiently small to believe that the model estimates the true value of the seizure rate modifier ψ with good accuracy. Within those small differences, it is noticeable that a smaller rate modifier $\psi_1 = 0.35$ is in general more identifiable to the model than the larger rate modifier $\psi_2 = 0.65$.

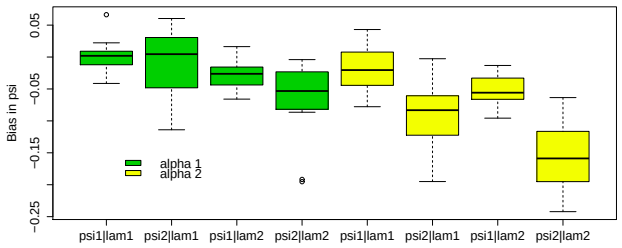
After obtaining small biases of the three parameters of interest, we would expect the residual study to reflect an acceptable goodness of fit of the model to the data set. In Figure 5.19 we show the Q-Q plots of the Cox-Snell residuals corresponding to Cowling's model. Observe that for both populations, the observed Cox-Snell residuals substantially fit their expected values, showing some dispersion for the late seizure times that does not diverge greatly from the center. Although the sub-population residuals shown in Figure 5.19a) tend to be slightly less variable from early seizures than those shown in Figure 5.19b), we notice that their deviance



(a) α estimates.



(b) $\lambda|\alpha$ estimates.

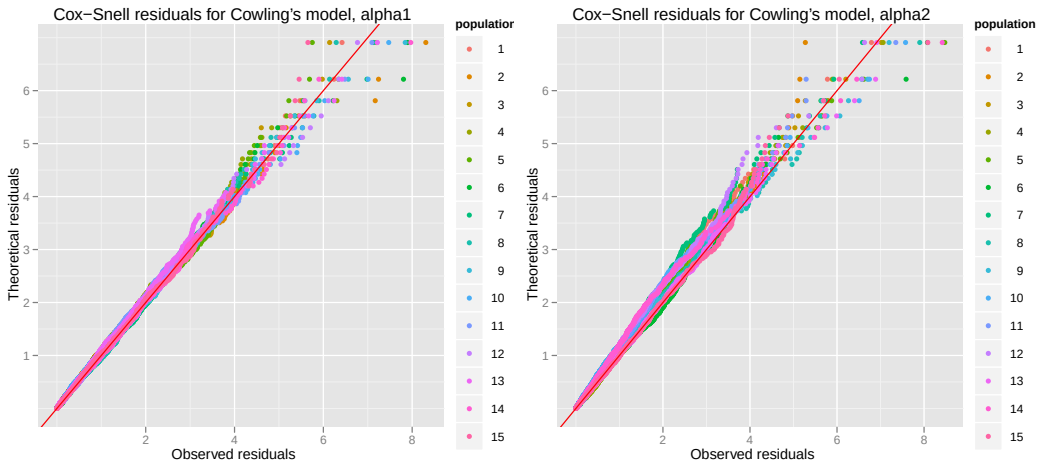


(c) $\psi|\alpha, \lambda$ estimates.

Figure 5.18: Box plots of the estimates' bias, for the three parameters of interest, when Cowling's model is fitted.

residuals, shown in Figure 5.20a), are scattered around -1, suggesting a deviation from the model of the same magnitude. When considering α_2, however, we have that although the Q-Q plot shows a slighter trend to underestimate than their counterparts for α_1, their corresponding deviance residuals, shown in Figure 5.20b), are centered around 0 and therefore suggest no significant deviation from the true model. Let us remember that, when fitting Cowling and Rogers models to the MESS data set, the estimated heterogeneity rate was found to be $\alpha = 1.99$ and $\alpha = 2.08$ for each of the respective models. This would suggest a homogeneity closer in behavior to the one assumed for the second sub-population of the simulations. Indeed, if we observe the Q-Q plots for the Cox-Snell residuals obtained in Figure 5.1 for Cowling's model, they appear to resemble the patterns produced in the corresponding Q-Q plots for the sub-populations in Figure 5.19b). The underestimation of the model

to the true seizure times may be an indication of a further covariate with influence that remains unaccounted for in the model, or even the indication that further modifications to the model are needed.



(a) $\alpha_1 = 0.5$ (b) $\alpha_2 = 4$

Figure 5.19: Q-Q plots of the Cox-Snell residuals for Cowling's model, the first graph corresponds to the population with α_1 and the second graph to the population with α_2.

5.6 Simulation studies for the zero-truncated joint model

In the following figures we observe a series of box-plots for the zero-truncated joint model under a simulation study. Let us recall that $\alpha_1 = 0.5$ and $\alpha_2 = 4$, hence the first value of α represents a high heterogeneity in the population, in contrast to the second value in which a smaller heterogeneity is assumed. We contrast the results predicted by the joint model always represented to the left of the figures, and the results presented by the truncated joint distribution, always shown at the right side of the figures. In order to keep the consistency with previous models under simulation, here we have 15 simulation runs for each value of the frailty term, with 1500 simulated values of pre-randomization counts and post-randomization times,

(a) $\alpha_1 = 0.5$ (b) $\alpha_2 = 4$

Figure 5.20: Plots of the deviance residuals for Cowling's model, the first graph corresponds to the population with $\alpha_1 = 0.5$ and the second graph to the population with $\alpha_2 = 4$.

under a Poisson-gamma mixture model.

The first great contrast between the findings is the observed bias for the estimates of α, shown in Figure 5.21, is that for the joint model the estimation bias lies around zero, with particularly narrow values for α_1, which is well represented for the truncated model as well. Meanwhile, when heterogeneity is small in the population, the estimation bias for the joint model is much larger, but for the truncated joint model the estimations are concentrated a unity of distance away. This might be one of the causes of why, even when the joint truncated model fits the estimates for $\lambda_1|\alpha_2$ and $\lambda_2|\alpha_2$ particularly well (Figure 5.22, to the right), the estimates for all values of ψ under α_2 deviate from the true value progressively (Figure 5.23), which also tends to happen for the values of ψ under α_2 in the joint model, but not so strongly.

In general, for the cases when the heterogeneity is high, both the joint model and the truncated joint model seem to fit well for α, λ and ψ.

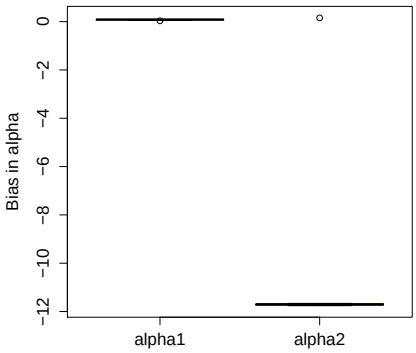


Figure 5.21: Cowling vs Truncated for α

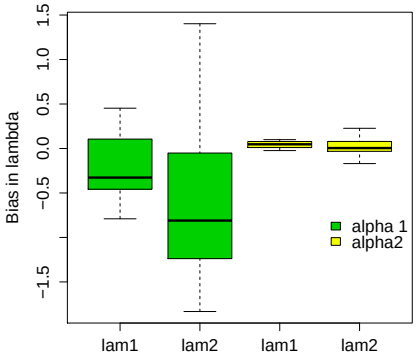


Figure 5.22: Cowling vs Truncated for λ

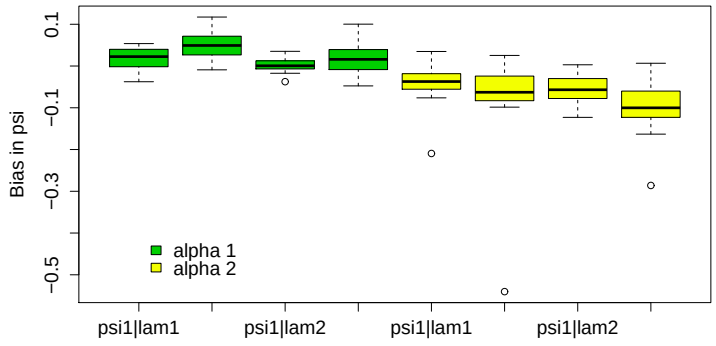
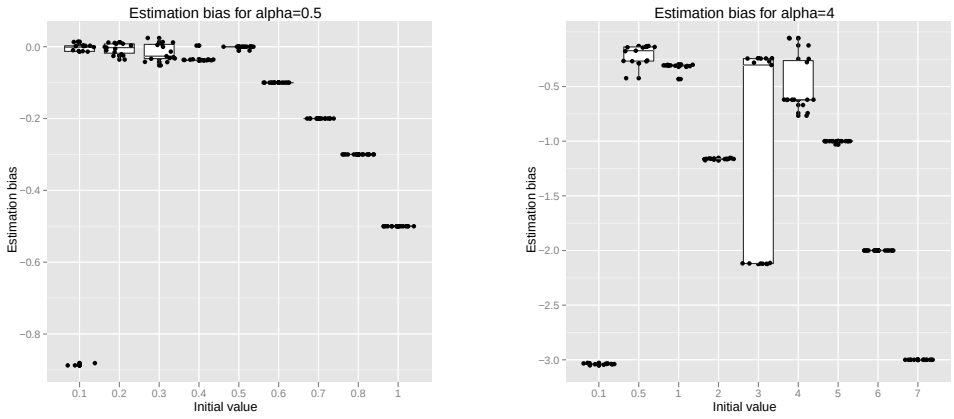


Figure 5.23: Cowling vs Truncated for ψ

Figure 5.24: Box-plots for the biases between the expected and the estimated values of α, λ and ψ.

As an attempt to understand the behavior of the truncated joint model when the heterogeneity of the population varies, we have performed a simulation study in which we only estimate the values of α. We proceed by simulating the occurrences of 1500 individuals as before, with the same values of α_1, α_2, λ_1, λ_2, ψ_1 and ψ_2. The difference lies in how we fit the truncated model. Instead of using a particular initial value α_0 of α for the process of optimization of the log-likelihood, we provide a series of 9 initial values of α_1 ranging from 0.1 to 1, and nine initial values for the estimation of α_2 ranging from 0.1 to 7. In Figure 5.25(a) we observe that, when the heterogeneity is high in the population, the truncated joint model tends to fit well for α even when the initial value is understated. For the case when the initial value is too small, the model concentrates the estimate either very close to the original value, or quite far away, as observed for the case in which $\alpha_0 = 0.1$. The bias however increases noticeably as the initial value surpasses the true value of the frailty term.

For the case when the original value of α is larger, depicted in Figure 5.25(b), we observe that the closest values of the parameter estimates are obtained, again, when the initial value is smaller than the actual value of the parameter. The general behavior of the estimates tend to be more erratic around the true value, but the same increasing trend of the bias is observed when the initial value surpasses the true value of α.



(a) Estimation bias for $\alpha = 0.5$.

(b) Estimation bias for $\alpha = 4$.

Figure 5.25: Estimation bias for simulated data under assumption of $\alpha = 0.5$ and $\alpha = 4$ respectively, when fitting a zero-truncated model with initial values for α shown in the horizontal axis.

We conclude by this simulation study that the joint truncated model tends to identify better the value of the frailty term when the heterogeneity in the population is large, and that the initial value provided for the optimization is of great consequence for the estimations.

5.7 Discussion of the residual analysis

In this chapter we were interested in performing a model checking study for four proposed models in epilepsy data. Our aim was to investigate the goodness of fit of the commonly used Cox proportional hazards survival model, along with two previously proposed accelerated-life survival models. For this purpose we selected Kim's model, which is a Cox model previously proposed by Kim et al [27] for the MESS data set, and another Cox model which considered not Kim's covariates, but Roger's covariates ([41]). Two accelerated-life models were considered from the works by Cowling [11] in his PhD thesis, and Rogers [39]. The first approach to study their goodness of fit to the MESS study was to introduce the theory and

methodology of residual data analysis. It was observed that, although the models did not appear to fit the data set well under the current assumptions, the residual plots provided insight on the outliers and extreme values found by the four models. Only Cox and Kim's models detected six outliers as such, and we refitted the four models taking out the information of such patients in order to study the difference in the model fits. Since we are not convinced the extreme values found by Cowling and Rogers models can yet be defined as outliers, we did not exclude the corresponding patients from the second study of the data set.

When adjusting the models to the reduced data set, we did find that Cox and Kim's models fits were improved to a certain degree, as would be expected. Cowling and Rogers models, however, improved only slightly and from the residual studies we observe that the models do not seem to fit the data very differently between the full and reduced data sets. Although this gives us signs that Rogers model in particular shows a greater stability over extreme values, the reduced data set did produce enough changes in the likelihood estimates to suggest that a simpler version of Cowling and Rogers models could be proposed, obviating some to the covariates considered previously. As expected, when six of the higher valued residuals were taken out, both of the latter models produced higher α estimates, suggesting a higher homogeneity in the data.

Some literature on residual analysis, particularly Collett [7], mention that there is little guidance as to how will the residual plots look like when the right model has been assumed for a data set. Given that in our previous studies we have observed four models being shown as inappropriate to the epilepsy data under study, we were interested in observing how would the plots change once we were certain of the right outcome. For this reason we simulated survival data consisting of exponentially distributed incidence times, and the number of incidences pre-randomization to be modeled by a Poisson distribution, which is almost the exact framework assumed for Cowling's model. The Cox model is deemed inappropriate

for this kind of data, and Cowling's model produces residual plots that demonstrate a much better goodness of fit. Observe that for these simulations we considered zero-truncated number of seizures, a fact that has not yet been taken into account for both Cowling and Rogers models. This is a natural generalization that we are interested in developing for further study, and it is currently under development. A second residual study that could be of interest but is not developed here, would consist of producing a series of simulations emulating the epilepsy seizure numbers and times as previously done, but additionally considering a cure fraction as Rogers' model assumes. It would be of interest to observe how would Cowling's model perform under this change for the residual analysis.

5.8 Lomax survival function and Kaplan-Meier curves

From the residual analysis methodology we have been able to observe and discuss several traits from the four survival models under consideration. The conclusions under such analysis, however, do not always point towards what might be the problem behind a residual plot with an apparent bad fit, and for this reason, we consider an alternative approach. Consider the joint and joint with cure fraction models; considering a relevant set of covariates (such as age, EEG outcome and type of epilepsy) we can obtain the corresponding covariate estimates for the seizure rate λ_i, the post-randomization seizure rate modifier ψ_i and the cure rate p_i for each individual i. Considering these covariates, we can find the corresponding survival function for the Lomax distribution, and compare its survival prediction under the models' fit for the MESS data, and contrast it to the observed Kaplan-Meier curves.

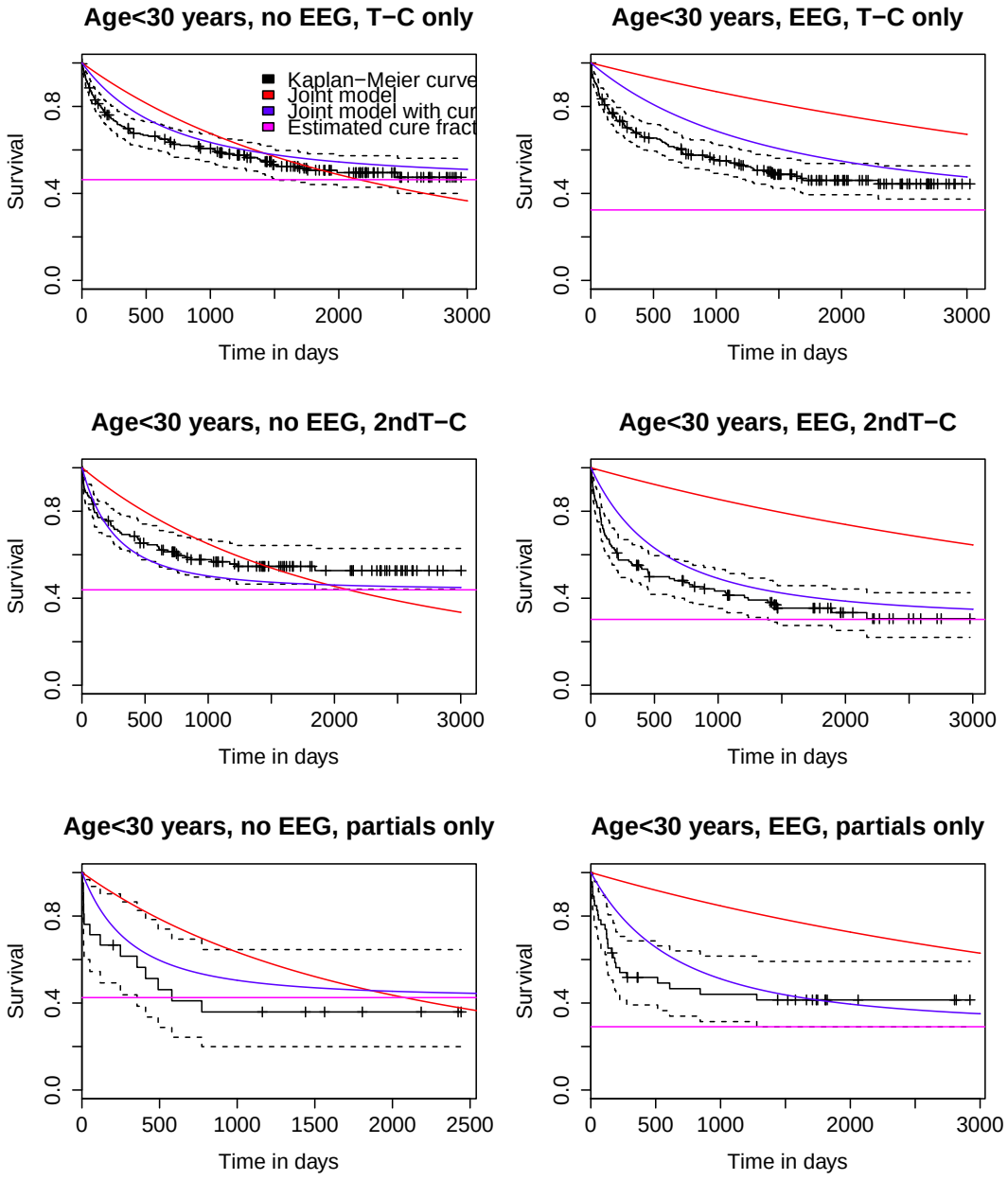


Figure 5.26: Kaplan-Meier curves for patients less than 30 years old, contrasted to survival functions from joint and joint with cure fraction models. The black line shows the Kaplan-Meier curve, the red and blue curves correspond to the survival curves corresponding to the joint and joint with cure fraction respectively. The purple line denotes the estimated cure fraction value.

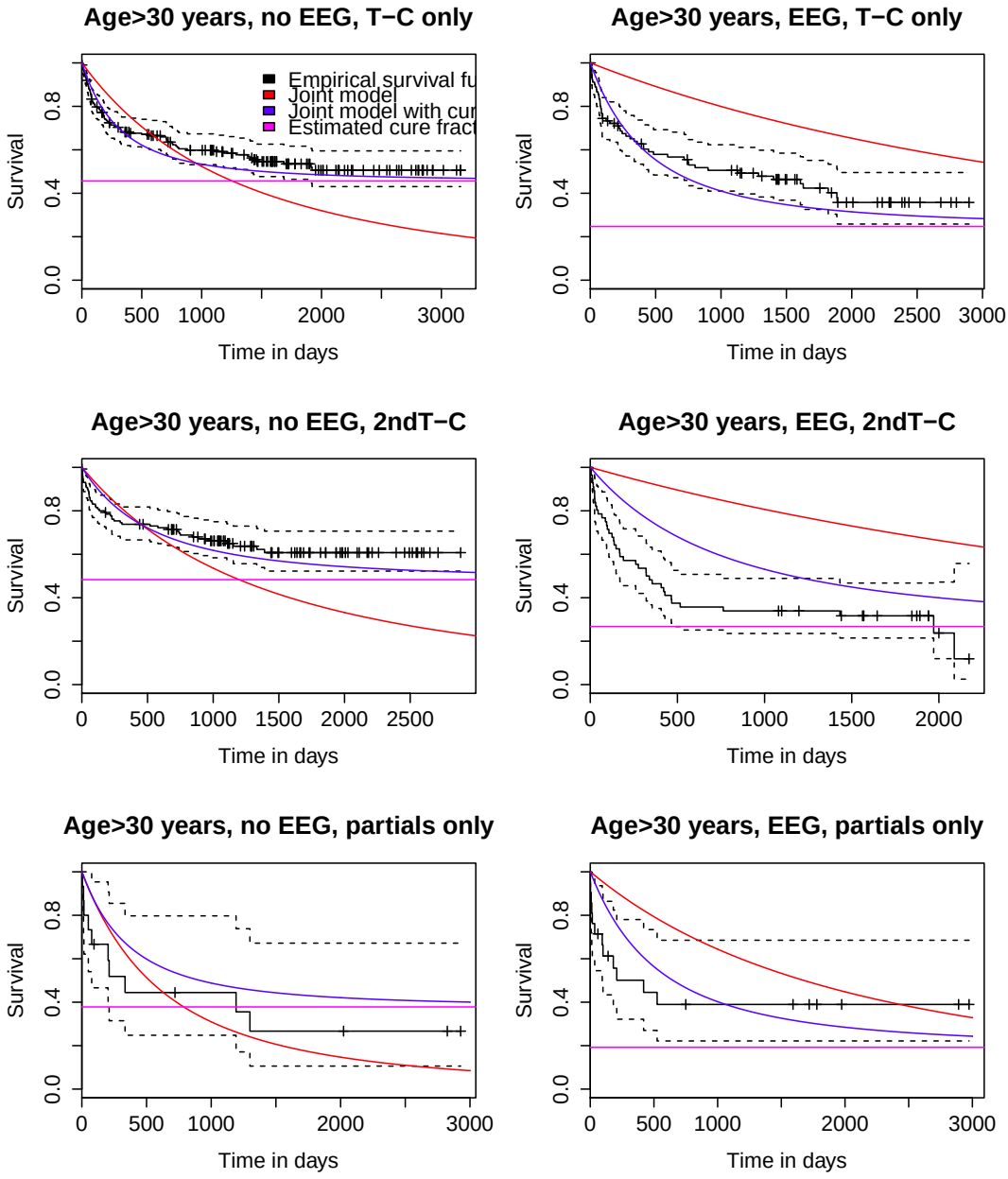


Figure 5.27: Kaplan-Meier curves for patients older than 30, contrasted to survival functions from joint and joint with cure fraction models. The black line shows the Kaplan-Meier curve, the red and blue curves correspond to the survival curves corresponding to the joint and joint with cure fraction respectively. The purple line denotes the estimated cure fraction value.

For the sets of plots in Figure 5.26 and Figure 5.27, we show the correspond-

ing contrasts for the population of patients younger and older than 30 years old, respectively. For each plot, we have partitioned the data for the combinations of presenting an EEG abnormality or not and the type of epilepsy. The Kaplan-Meier curve, shown in black, corresponds to the partition mentioned in the title. The red curve corresponds to the Lomax survival curve under the estimated parameters for that subpopulation, and similarly the blue curve corresponds to the joint model with a cure fraction, which is the estimated proportion of the population that experiences remission and is signaled by the purple line.

We observe that under these survival function predictions, the joint model tends to overestimate the survival of individuals showing an abnormal EEG, regardless of their age group. Although the joint model with cure fraction also tends to overestimate the survival curve for patients younger than 30 years old and presenting abnormal EEG outcomes, compared to the joint model it overestimates the survival curve in a lesser measure, and the steep descent for early seizures provides a more desirable fit. Observe that the predicted curves correspond to a mean age for each respective age group, and that the corresponding age quartiles for this model are plausible to contain the observed Kaplan-Meier curves.

For the case of patients younger than 30 years old and presenting a normal EEG outcome, the joint model survival function approaches the Kaplan-Meier curve in a much better way, although the early seizure steep descent is always overestimated by the model. Precisely in these cases is that the joint model with a cure fraction approaches the observed survival curves, specially for the Tonic-Clonic seizures only, almost accurately. This level of a good fit is observed again for the joint model with cure fraction for patients older than 30 years old and with normal EEG. In these cases the model lies close to the observed survival curve, whilst the corresponding predicted survival function for the joint model consistently underestimates the Kaplan-Meier curves. A particular case in which the joint model with cure fraction appears to overestimate the observed survival curve is when patients

are older than 30 years old, present an abnormal EEG and second generalized Tonic-Clonic seizures. For this case, it would be of particular interest to plot the respective predicted survival curves under the upper and lower quartiles of the age group.

For the following Figure 5.28 and Figure 5.29 the contrasts of survival function predictions are shown for the population randomized to delayed or immediate treatment respectively. As shown before, in these plots we partition the epilepsy data for the combinations of EEG outcome and epilepsy type. Note that in these cases the data is not partitioned by age at randomization.

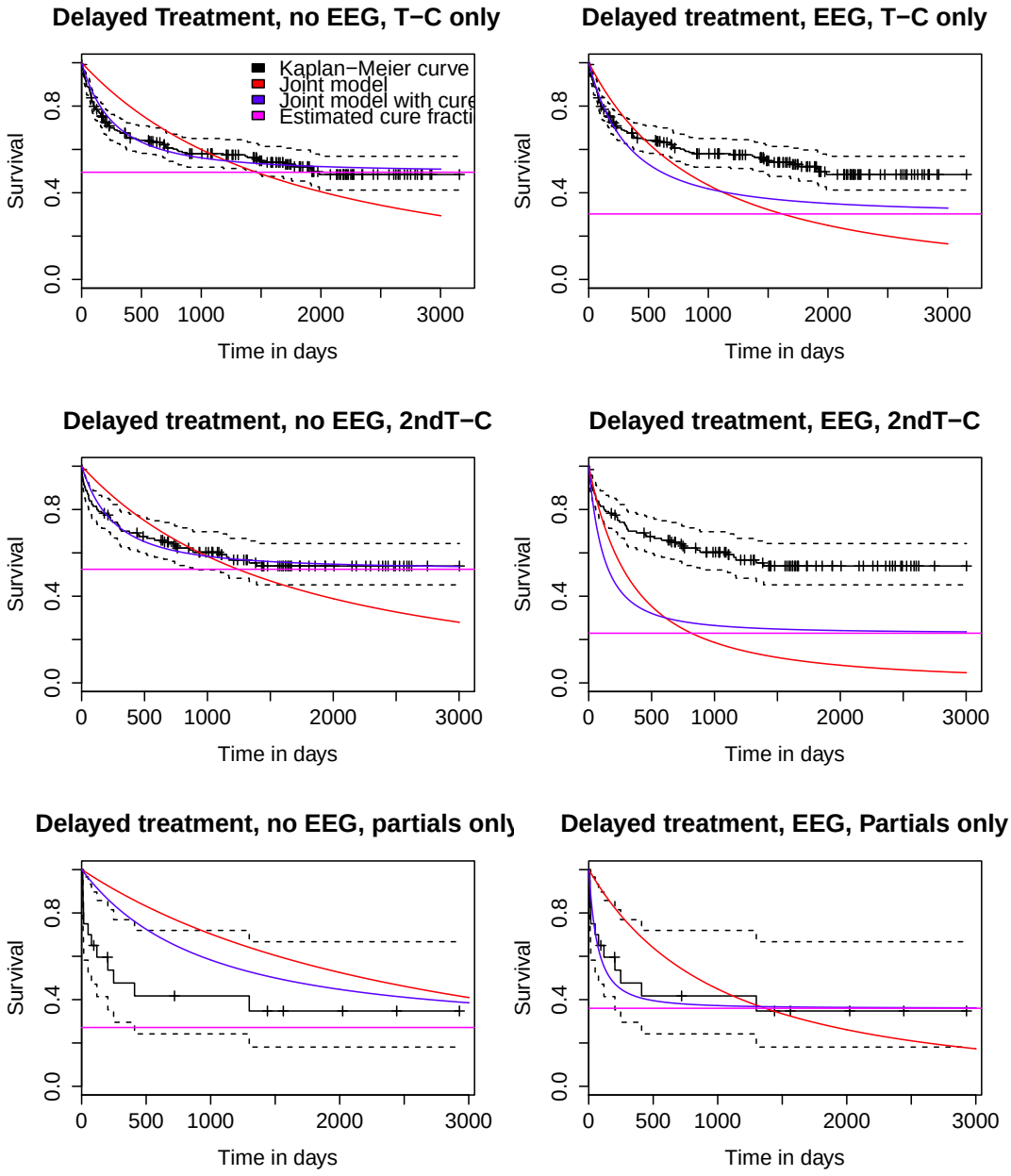


Figure 5.28: Kaplan-Meier curves for patients randomized to delayed treatment, contrasted to survival functions from joint and joint with cure fraction models.

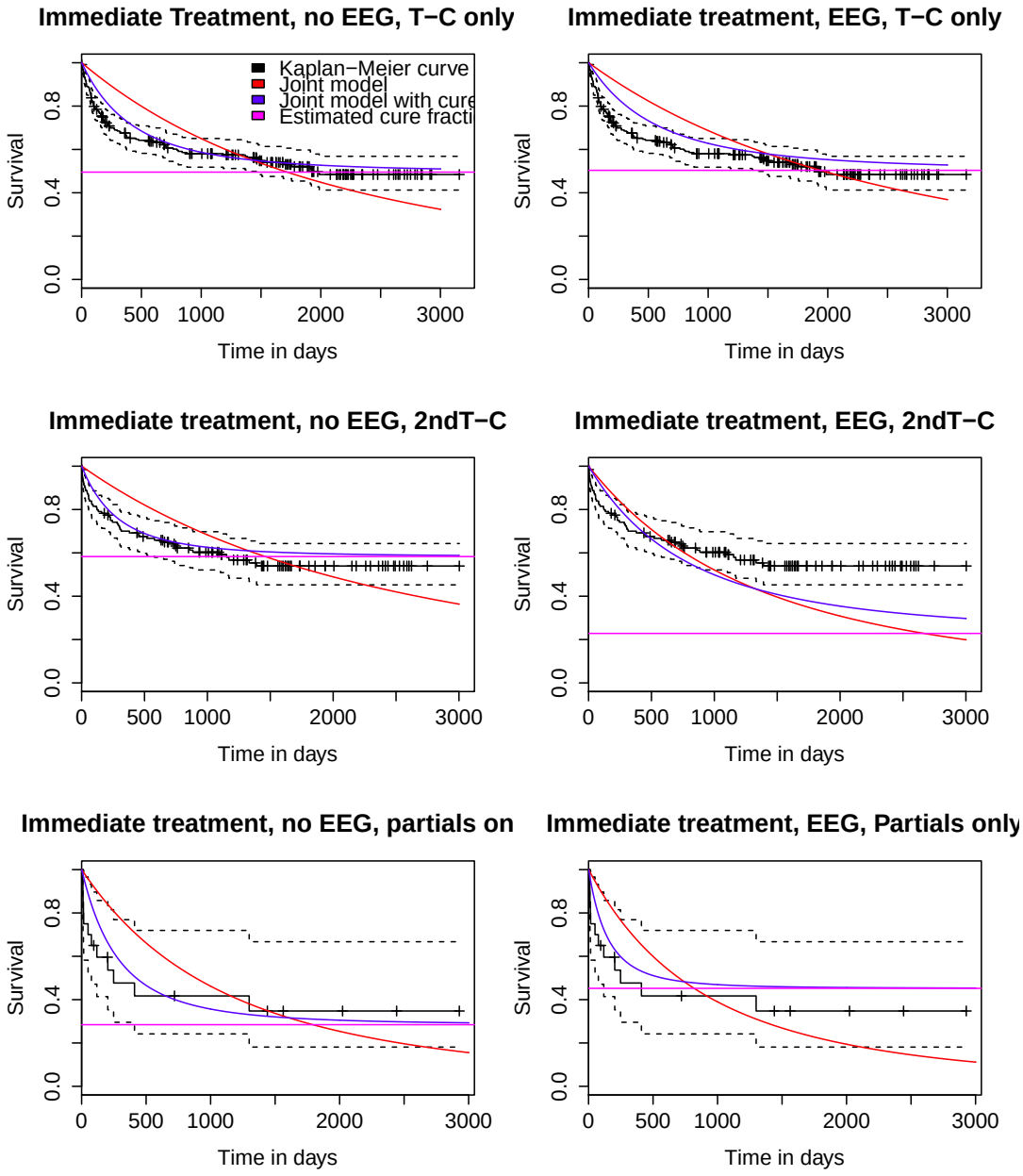


Figure 5.29: Kaplan-Meier curves for patients randomized to immediate treatment, contrasted to survival functions from joint and joint with cure fraction models.

In contrast to the plots in which the populations are partitioned by age at randomization, when we partition according to the treatment allocation the joint model tends to over and under estimate the Kaplan-Meier curve in a smaller mea-

sure. Indeed both models seem to fit the population survival better, except particularly in the cases in which patients are randomized to delayed treatment, present an abnormal EEG and second generalized tonic-clonic seizures or tonic-clonic seizures only. In this case the joint model predicts a much steeper descent of the survival curve, and even the joint model with the cure fraction overestimates the hazard in the population, since the model estimates a much lower proportion of patients attaining remission than the actual number of patients not presenting a seizure post-randomization. This overestimation of the hazard in the population is observed again for both models, but in a more controlled manner, when the patients present abnormal EEG and second generalized tonic-clonic seizures, but have been randomized to immediate treatment. From these plots we observe that the discrepancies between observed and predicted survival curves seem to be due to the lack of fit of the cure fraction estimate. Possible remedies could be found by finding the covariates that are most descriptive of the proportion of patients who do not experience a seizure post-randomization, and will be considered as part of the further work.

Chapter 6

Conclusions and Discussion

Usually medical epilepsy studies include baseline data of the patients, comprising the patient's history previous to the treatment. Either for convenience or because of familiarity, Cox's proportional hazards models tend to be commonly used for medical recurrence data, which usually considers the patient's historical additional information as a fixed covariate. It has been argued that these models tend to provide a deficient fit to medical recurrence data, and for the case of epilepsy specially. In Kwong & Hutton[30], whilst investigating the proportionality of parameters under model misspecification (as proposed by Hutton & Solomon[25]) for two data sets, found that for the epilepsy data, this condition does not hold even when the regression coefficients are assumed small. The AIC measure for the models they considered showed that the accelerated life models fitted the data better than the proportional hazards model. Indeed, the AIC statistic indicated a better fit under the log-normal, log-logistic and gamma distributions, whilst the Weibull, being both a proportional hazards and an accelerated life model distribution provided the worse fit. The reason the results from Hutton & Solomon do not seem to follow is due to the high skewness of the hazard function, for which the center of density concentrates near zero. In our case, the MESS epilepsy data also shows from the Kaplan Meier curves a high hazard skewness due to a large number of early seizures.

This observation and the result from Hutton & Monaghan[24], which showed that accelerated life models are more robust under misspecification, suggest that, as in Kwong & Hutton, we consider an accelerated life model instead of a proportional hazards model.

Previous studies have proposed a more informative model for recurrent events such as epilepsy seizures. The Poisson distribution has been proposed to account for the overdispersion in the data (Hougaard et al.[23]), joint models have been developed to account for the pre-randomization and post-randomization recurrence counts (Marshall & Olkin[34]), and bivariate semi-parametric models have been presented in an attempt to account for the recurrent event times and counts jointly (Cook & Wei[9]). In the works by Cowling[10] and Rogers[39] these concepts have been employed in the form of a Poisson-Gamma mixture model, where the seizure recurrences can be thought to have an underlying point process to describe the failure recurrence. Indeed, the model assumes that, conditional on a gamma distributed frailty term, the epileptic seizures can be modelled as a Poisson distributed number of occurrences, in exponentially distributed recurrence times. Since the underlying seizure rate is assumed to be the same for both the Poisson and exponential distributions, these models propose a joint model which links the seizure counts and times jointly, but estimates a post-randomization effect that is assumed to have a multiplicative effect on the seizure recurrence rate. One important improvement from previous models consists in considering an individual seizure rate for each patient, and an individual treatment effect on such a rate. This assumption allows the model to account for a more informative prognosis for each patient, and a more detailed prognosis for the population by considering the variation between individuals. From the Kaplan-Meier curves shown in Chapter 3 (Figures 3.1 and 3.2) it has been observed that there is a substantial proportion of the population that, after randomization, do not experience another seizure. This phenomena is included in the model by Rogers by proposing a cure fraction of the population, such that a set

of covariates is assumed to have a logistic relationship to the proportion of patients who do not present a new seizure from randomization.

In this thesis we study the contrast between the commonly accepted Cox proportional hazards models and the models proposed by Cowling and Rogers, denominated here as the joint and joint with cure fraction models respectively. The Cox models were presented under two covariate sets, mainly the covariates used in the paper by Kim et al.[27] (neurological disorder, EEG outcome and the logarithm of the number of seizures pre-randomization), and the covariates used for the Cowling and Rogers models (age at randomization, epilepsy type, treatment allocation, EEG outcome and interactions). A comparison between their coefficient parameter estimates (Tables 3.4 and 3.5) show that, under Kim et al.'s original set of covariates, all factors appear to be significant under the Cox and the Rogers models, whilst for the Cowling model the neurological disorder outcome is only significant for the pre-randomization seizure rate. The joint model with a cure fraction shows an improvement of the fit from its larger log-likelihood value, but due to the scarce amount of information of the neurological disorder covariate for this epilepsy study, we propose a second set of covariates which does not include it. Under this new set of covariates we compare the joint models with the Cox model, shown in Table 3.6, and we find that a significant difference between their predictions lies in the relevance of the seizure type. Whilst the joint models show that the two types of epilepsy have a relevant contribution to the seizure rates, consistently for pre and post-randomization rates, whilst in contrast, the Cox model only consider their interactions with the treatment allocation and the EEG outcome to be of no significance. The EEG outcome factor contributes a significant change only under the joint model with a cure fraction, but the EEG interaction with the 2nd. tonic-clonic seizure provides a non-significant contribution only for this same model. The number of seizures pre-randomization is a very relevant covariate for the Cox model, whilst the estimated value of κ translates to 50.10% of the population not experienc-

ing a seizure post-randomization. The difference between log-likelihoods between the two joint models shows an improvement of the model fit for the data under the assumption of the existence of a cure fraction of the population.

6.1 Summary of Thesis

In this thesis we contrast the goodness of fit of models that have been proposed in the survival analysis framework, producing an extensive residual analysis for proportional and accelerated life models for epilepsy data. This thesis proceeds to propose a number of models which consider a joint distribution for the number of seizure counts as well as the first time of occurrence post-randomization, considering a gamma distributed mixture distribution to account for overdispersion, and a zero-truncation to account for the known fact that patients must have at least two unprovoked seizures to be diagnosed with epilepsy. An alternative model studies the effect of the overdispersion parameter α depending on covariates, and suggests that for further work it is desirable to implement this covariate dependent frailty term for the zero-truncated model.

In Chapter 2 we describe the epilepsy fundamental medical definitions, as well as the nature of the object of our study, the MESS data set. We provide a brief introduction to likelihood estimation and finally, to survival analysis theory for censored data. In Chapter 3 presents an extensive discussion of previous models proposed for medical recurrent events, guiding the reader chronologically from semi-parametrical univariate models, to the models which constitute our initial interest of study. We discuss the reasons that lead us to contrast proportional hazards models to accelerated life models.

In Chapter 4 we propose three models for recurrent events, particularly those in which the patient is known to have experienced the event of interest at least once in the past. These models consider a joint distribution where the number of oc-

currences before randomization are considered to be conditionally distributed as a Poisson with intensity λ for each patient, and the post-randomization seizures are assumed to follow an exponential distribution, conditional on a gamma distribution as the mixture frailty term to account for the overdispersion in the population. The case in which the zero-truncated model considers a fraction of the population to be seizure-free post-randomization is still under work, but the joint distribution, the likelihood and its respective derivatives are shown here. A final model is proposed which generalizes a Poisson-Gamma mixture model, by considering the gamma parameter α to depend on covariates. This generalization is shown to perform better than the original model, according to the difference between their log-likelihood values.

Finally in Chapter 5 we contrast the performance of four models for the MESS epilepsy data. These four models have been proposed before in the papers by Marson et al.[35], Kim et al.[27], Cowling et al.[11],[10] and Rogers et al.[41],[39]. The residual analysis showed that Cox models do not fit well the patients in the data, except for those producing low-valued residuals. The identification of outliers is discernable however, which contrasts from the joint models, where the existence of outliers is not clear. The joint models' fits also present a departure from the data observations, in different ways. The joint model proposed by Cowling et al. produces overestimated residuals, but the pattern is consistent along an alternative re-scaled line. The model proposed by Rogers et al. seems to fit well more than half the population, but the remaining residuals deviate from the equality line, showing a pattern that could possibly be due to a missing covariate from the analysis.

6.2 Truncated and Covariate-Dependent Frailty Term Model Conclusions

Aiming to propose a more accurate model for epilepsy, where the patient is recorded to have at least one seizure, we presented and implemented the zero-truncated joint model. When fitted to the MESS epilepsy data, the truncated model provides a larger log-likelihood value than the joint and the joint with cure fraction models. Judging by the model fits to the partitioned data by epilepsy type, we find that the truncated model performs best for the population of patients with Tonic-Clonic seizures only, with the worst performance being provided for the patients with partial seizures only, in which the estimated heterogeneity in the population is greater. From this it would appear that the truncated model performs better when the population is homogeneous, but this is likely to be more a correlation between the covariates, given that we know that patients with partial seizures are more heterogeneous in symptom and seizure frequency than patients with generalized seizures. An indication of such correlation between parameters is given by the box-plots in Figure 5.18, where subject to simulation, we observe that the estimates of λ and ψ are more or less biased, depending on the value of α.

The zero truncated model does not consider patients with no seizures pre-randomization, and hence concentrates its mass towards positive values and considers a lower heterogeneity between the individuals in the study. The value of the frailty term parameter, α, is estimated to be five times larger under the truncated model compared to the joint and joint with cure models, predicting a much larger homogeneity in the population. Another important difference between the models resides in the fact that the age of the patients at randomization becomes significant only under the truncated model, for patients with generalized and second generalized seizures, as can be observed in the table for the partitioned data according to epilepsy type. For patients with partial seizures only, however, the age covariate

becomes nonsignificant for this model.

From the model considering the overdispersion parameter as dependent on covariates, we observe a consistent estimation of the homogeneity prediction with the corresponding predictions given by the zero-truncated model. When considering α as dependent on the type of epilepsy, the patients with partial and Tonic-Clonic seizures only are deemed to have the highest and the lowest heterogeneity estimates respectively. When α is considered to be dependent on the EEG outcome, the model reflects the known similarities and dissimilarities between patients with different EEG outcomes, i.e. that the patients with a normal EEG will be more homogeneous than those who have an abnormal EEG outcome. For both α covariate dependencies we find that the pre-randomization covariate estimates do not present a significant change from Cowling's model. For the post-randomization covariate estimates, the important differences are observed for the EEG outcome, where its pure effect is deemed negligible compared to the standard deviation size. The EEG's interactions with treatment and the second generalized Tonic-Clonic seizures remain significant, as opposed to the EEG and Tonic-Clonic interaction which remains non statistically significant.

The zero-truncated model with cure fraction presents no significant prognostic difference from the truncated model without the cure fraction. Furthermore, it suggests the worst fit to the MESS data among the joint models under consideration, when compared by log-likelihood value. However, due to the nature of some quotients in the likelihood and the derivatives, the values of the log-likelihood function tend to be greatly affected by the high number of seizure counts for some of the patients. Indeed, when the five patients with the largest number of pre-randomization counts are omitted from the analysis, the estimated level of homogeneity in the population, indicated by the values of α decreases, since α in turn decreases from 7.997 to 7.441 and the log-likelihood value increases from -8020.696 to -7955.125. Further investigation of the model's behaviour is left as a future work, from the

observation of its parameter relationships between one another and possible future residual analysis.

6.3 Goodness of Fit Conclusions

The Kaplan-Meier survival curves for the MESS data set (Figures 3.1 and 3.2) show the first, second and fifth seizure occurrences for the population. The progression of the survival curves either stratified by drug or by treatment allocation cast a reasonable doubt about the assumption of a Cox models as the underlying recurrence model, since the proportionality assumption does not seem to hold between the subpopulation hazards. In an attempt to understand better the fit performance between the Cox and the joint models proposed by B. Cowling and J. Rogers, we provide a brief introduction to residual analysis and proceed to obtain four types of residuals of the data under two Cox models and Cowling's and Rogers' models.

The Cox-Snell residuals plotted in the exponential Q-Q plots show that the joint model does not fit the data accurately, but rather, that the obtained residuals from the model are overestimating the exponential residuals by a factor of two, but otherwise the residuals follow this trend consistently and therefore the reason for this overestimation would be a natural object of study for this model. Furthermore, there is no evidence that the censoring nor the type of epileptic seizure produce a trend in these residual plots. Contrasting with the consistent trend of the residuals for the joint model we obtain the exponential Q-Q plots for the two Cox models and the joint model with a cure fraction, where we observe that for all these cases the models fit well different proportions of patients with early seizures, and then diverge into an overestimating trend of a factor of three for the Cox models, and for the joint model with cure fraction, with a sinusoidal ascending pattern. Both Cox model's Cox-Snell residuals present a very clear departure from the identity line, which we expected to observe due to the lack of proportionality of the hazard

functions observed in the Kaplan-Meier plots, and from the limitation of the model to use the seizure count as a variable. We are more interested in discerning the reason behind the residuals behavior observed under the joint models proposed by Cowling and Rogers.

For the Cox models there are a four individuals that we consider as outliers and take out of the study. The resulting new Cox-Snell residuals show a slightly less overestimated fit of the Cox models, which gives evidence that the omittance of high-valued times to seizure results in a better fit of the Cox models. This same overestimation decrease is observed in the residual Q-Q plots for the joint models proposed by Cowling and Rogers when omitting the four patients with the greatest residual values in the population. In the case of the joint models, the patients corresponding to these highest-valued residuals present remarkably high seizure times, they are all young adult females allocated to immediate treatment, present abnormal EEG, few pre-randomization seizures and second generalized tonic-clonic seizures. In the model we have accounted for all these covariates, and in the model fit for the population omitting these patients we observe no significant change in the underlying seizure rate λ_i or the homogeneity rate α, but in the post-randomization rate modifier ψ_i. Since some of the covariates such as the Tonic-Clonic seizure only and interactions decrease in significance, a model with a reduced number of covariates could be an improvement for the model fit. From these patients, three of them belong to the same clinical centre, a covariate we have not yet considered for the study, since we need a deeper analysis of the relationship and possible correlation between the seizure time reported and the clinical centre. Because of time constraints, we have not performed a correction measure for this covariate correlations, but remains as a desirable project for future work.

From the martingale and deviance residuals we find that there is no noticeable trend due to epilepsy type or treatment allocation, and support the outlier findings observed in the Cox-Snell residuals for the Cox models. From the various residual

plots, particularly when plotting against the age at randomization, we observe that only the joint model considering a cure fraction shows the existence of trends of residuals, although not corresponding to the age covariate. We observe that the residuals seem to group around four lines or levels, which seem to correspond to the levels observed in the exponential Q-Q plot. This behavior could be due to a covariate that is not being considered in the model, and that with the existence of the cure fraction becomes relevant. However, the lack of information about how a residual plot would look like when the model fit is good, we performed a residual analysis under simulated data. We simulated a Poisson process with different combinations of parameters, and from the residual plots concluded that under the Cox models, the exponential Q-Q plots do resemble the corresponding plots under the MESS epilepsy data. Under the joint model, however, the exponential Q-Q plots show that the residuals accumulate around the identity line, and the boxplots of the estimates' bias show that the biases are small for all three parameters of interest, where the larger the heterogeneity in the population, the more accurate is the model in estimating α and ψ_i, and the less accurate it is in estimating λ_i.

As an alternative goodness of fit analysis, we produce plots contrasting the Kaplan-Meier curves against the survival functions predicted from the Lomax and Lomax with cure fraction models, which correspond to the Cowling and Rogers models respectively. These plots under several subpopulations show that for patients under 30 years of age and presenting a normal EEG, the Lomax survival function approximates the Kaplan-Meier curve, whilst for the remaining cases it either over or underestimates the survival of the population. When introducing the cure fraction the Lomax survival function approaches the observed survival of the population much more accurately, dependent, however, in the ability of the Rogers model to estimate the cure fraction in the population. Both age at randomization and the combination of EEG outcome and epilepsy type appear to be significant covariates that should be considered for the cure fraction. These plots provide evidence that

the inclusion of the cure fraction in a Lomax distribution provides a good model fit for the seizure recurrence times.

6.4 Future Work

From the residual analysis we observed that the joint model which considers the frailty term as dependent on the epilepsy type controls the behavior of the residuals, it is of particular interest to analyze further this model under a more clinically selected set of covariates, as well as implement a joint model with a cure fraction with a covariate dependent frailty term. We would expect that by considering a model where both the frailty term and the cure fraction are estimated through a set of covariates will provide a more accurate fit of predicted survival curves, when contrasted to the Kaplan-Meier curves. From Figures 5.26 through 5.29 there is evidence that covariates such as age at randomization, EEG outcome and epilepsy type are particularly significant for the cure fraction.

Since the residual analysis was carried out before the zero-truncated models were developed, it is of special interest to implement the goodness of fit analyses presented in Chapter 5 for this model. In order to obtain the survival function we will need to approximate numerically the cumulative distribution of the zero-truncated function. This may be done by finding the integral of the density function numerically, or by considering the intractable integral in the expression (4.7) as the -1 moment of a Lomax distribution.

Here we have presented three joint models that derived from the models by Cowling and Rogers, and we aim as a part of possible future work, to contrast these models to a competing-risks survival analysis, in which seizure occurrence, death, drop-out and remission, could be considered as the risks of interest and competing with each other.

Chapter 7

References

Bibliography

[1] Aalen, O., Andersen, P., Borgan, O., Gill, R., Keiding, N.(2010). History of applications of martingales in survival analysis. *Journal Electronique d'Histoire des Probabilits et de la Statistique*, **5(1)**, *pp.*1-28.

[2] Andersen, P. K., Borgan, O., Hjort, N. L., Arjas, E., Stene, J. & Aalen, O.(1985). Counting Process Models for Life History Data: A Review (With Discussion and Reply). *Scandinavian Journal of Statistics*, **12**, 2, *pp.*97-158.

[3] Appleton, R. and Marson, A.(2009). Epilepsy. *Oxford University Press*, Third Edition.

[4] Bonnett, L., Smith, C. T., Smith, D., Williamson, P., Chadwick, D. & Marson, A. G.(2012). Prognostic factors for time to treatment failure and time to 12 months of remission for patients with focal epilepsy: post-hoc, subgroup analyses of data from the SANAD trial. *Lancet Neurology*, **4**, *pp.*331-40.

[5] Cameron, A.C. & Trivedi P. K.(1998). Regression Analysis of Count Data. *Cambridge University Press*, First Edition.

[6] Cohen, A. Clifford(1991). Truncated and censored samples: theory and applications. *New York : Dekker*, First Edition.

[7] Collett, D.(1994). Modelling Survival Data in Medical Research. *Chapman & Hall*, First Edition.

[8] Cook, R. J. & Lawless, J. F.(2002). Analysis of Repeated Events. *Statistical Methods in Medical Research*, **11**,*pp.*141-166.

[9] Cook, R. J. & Wei, W.(2003). Conditional analysis of mixed Poisson processes with baseline counts: implications for trial design and analysis. *Biostatistics*, **4**, 3, *pp.*479-494.

[10] Cowling, Benjamin. J.(2003). Thesis: Survival Models for Censored Point Processes. *Thesis for the Degree of Doctor in Philosophy*, University of Warwick, Department of Statistics.

[11] Cowling, B. J., Hutton, J. L. & Shaw, J. E. H.(2006). Joint modelling of event counts and survival times. *Applied Statistics*, **55**, *Part* 1, *pp.*31-39.

[12] Cox, D. R.(1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, **B**, *34*, *pp.*187-220.

[13] Crowder, Martin(2001). Classical Competing Risks. *Chapman & Hall*, First Edition.

[14] Crowley, J. & Hu, M.(1977). Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association*, **72**, 357, *pp.*27-36.

[15] Engel Jr., Jerome(2006). ILAE classification of epilepsy syndromes. *Epilepsy Research - Elsevier*, **70, Suplement**, *pp.*5-10.

[16] Farewell, Daniel & Henderson, Robin(2010). Longitudinal Perspectives on Event History Analysis. *Lifetime Data Analysis*, **16**, 1, *pp.*102-117.

[17] Farrington, C. P.(2000). Residuals for Proportional Hazards Models with Interval-Censored Survival Data. *Biometrics*, **56**, 2, *pp.*473-482.

[18] Gill, Richard D.(1984). Understanding Cox's Regression Model: A Martingale Approach. *Journal of American Statistical Association*, **79**, 386, *pp.*441-447.

[19] Greenwood, M. & Yule,G. U.(1920). An Inquiry into the Nature of Frequency Distributions of Multiple Happenings, with Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents. *Journal of the Royal Statistical Society*, **83**, *pp.*255-279.

[20] Hastings, N. A. J. & Peacock, J. B.(1974). Statistical Distributions. *Butterworths*, First Edition.

[21] Hougaard, P.(1986). A class of multivariate failure time distributions. *Biometrika*, **73**, 3, *pp.*671-678.

[22] Hougaard, P.(2000). Analysis of Multivariate Survival Data. *Springer*, First Edition.

[23] Hougaard, P., Ting Lee, M. & Whitmore, G. A.(1997). Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes. *Biometrics*, **53**, 4, *pp.*1225-1238.

[24] Hutton, J. L., Monaghan, P. F.(2002). Choice of parametric Accelerated Life and Proportional Hazards Models for Survival Data: Asymptotic results. *Life Data Analysis*, **8**,*pp.*375-393.

[25] Hutton, J. L., Solomon, P. J.(1997). Orthogonality in Mixed Regression Models for Survival Data. *Journal of Royal Statistical Society*, **B**, **59**, No.1, *pp.*125-136.

[26] Kampstra, Peter(2008). Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. *Journal of Statistical Software, Code Snippets*, **28**, *pp.*1-9.

[27] Kim, L. G., Johnson, T. L., Marson, A. G., Chadwick, D. W.(2006). Prediction of risk of seizure recurrence after a single seizure and early epilepsy: further results from the MESS trial. *Lancet*, **5**, *pp.*317-22.

[28] Klein, Jhon P. & Moeschberger,Melvin L.(2003). Survival Analysis, Techniques for Censored and Truncated Data. *Springer*, Second Edition.

[29] Kumar Nayak, Tapan(1987). Multivariate Lomax Distribution: Properties and Usefulness in Reliability Theory. *Journal of Applied Probability*, **24**, *pp.*170-177.

[30] Kwong, G. P. S., Hutton,J. L.(2003). Choice of parametric models in survival analysis: applications to monotherapy for epilepsy and cerebral palsy. *Applied Statistics*, **52**, *Part 2*, *pp.* 153-168.

[31] Lawless, Jerald. F.(2002). Statistical Models and Methods for Lifetime Data. *Wiley-Interscience*, Second Edition.

[32] Leemis, Lawrence M.(1995). Reliability, Probabilistic Models and Satistical Methods. *Prentice-Hall*, Second Edition.

[33] Lindeboom, Maarten & Van Den Berg, Gerard J.(1994). Heterogeneity in Models for Bivariate Survival: The Importance of the Mixing Distribution. *Journal of the Royal Statistical Society, Series B (Methodological)*, **56**, 1, *pp.*49-60.

[34] Marshall, A. W. & Olkin, I.(1990). Multivariate Distributions Generated from Mixtures of Convolution and Product Families. *Topics in Statistical Dependence*, **16**,*pp.*371-393.

[35] Marson, A., Jacoby, A., Johnson, A., Kim, L., Gamble, C. & Chadwick, D.(2005). Immediate versus deferred antiepileptic drug treatment for early epilepsy and single seizures: a randomised controlled trial. *Lancet*, **365**, *pp.*2007-13.

[36] McCullagh, P. & Nelder, J. A.(1989). Generalized linear models. *London : Chapman and Hall*, Second edition.

[37] Meeker,W. Q. & Escobar, L. A.(1998). Statistical Methods for Reliability Data. *Wiley Series in Probability and Statistics*, First Edition.

[38] Mikosch, Thomas(2009). Non-life Insurance Mathematics, An introduction with the Poisson Process. *Springer*, 2nd. Edition.

[39] Rogers, Jennifer K.(2011). Thesis: Statistical Models for Censored Point Processes with Cure Fractions. *Thesis for the Degree of Doctor in Philosophy*, University of Warwick, Department of Statistics.

[40] Rogers, J. K., Hutton, J. L.(2012). Joint modelling of pre-randomization event counts and multiple randomization survival times with cure rates: application to data for early epilepsy and single seizures. *Journal of Applied Statistics*, **3**, *pp.*546-562.

[41] Rogers, J. K., Hutton, J. L., Marson, A. G. & Chadwick, D. W.(2012). Assessing the risk of subsequent tonic-clonic seizures in patients with a history of simple or complex partial seizures. *Journal of Neurology, Neurosurgery and Psychiatry*, **83**, *pp.*803-809.

[42] Smith, Peter J.(2002). Analysis of Failure and Survival Data. *Chapman & Hall/CRC*, First Edition. *StLouis* St Louis, Erik K., Ficker, David M. and O'Brien, Terence J., Editors(2015). Epilepsy and the Interictal State: Comorbidities and Quality of Life. *John Wiley & Sons*, First Edition.

[43] Therneau, Terry M., Grambsch, Patricia M. & Fleming, Thomas R.(1990). Martingale-based Residuals for Survival Models. *Biometrika*, **77**, 1, *pp.*147-160.

[44] Tobias, Paul A. & Trindade,David C.(2011). Applied Reliability. *CRC Press*, Third Edition.

[45] Venables, W. N., Ripley, B. D.(1999). Modern Applied Statistics with S-PLUS. *Springer*, Third Edition.

[46] Wang, Lee(2003). Statistical Methods for Survival Data Analysis. *Wiley Interscience*, 3rd. Edition.

Appendix A

Parameter Orthogonalization

The importance of parameter orthogonalization, and the means to obtain orthogonality, were developed and discussed in the paper by Cox & Reid[12], in 1987. The main goal of this paper is to address parametric probability models, which in turn are in term of a set of parameters θ, consisting of a subset of parameters of interest θ_1, and a subset of nuisance parameters θ_2. Although it is mentioned that both sets of parameters could be orthogonalized from each other, the paper focuses in the direct problem when θ_1 is a scalar, and θ_2 is the set of remaining nuisance parameters.

Previously, the maximum likelihood estimator (MLE) $\hat{\theta}_2$ of θ_2 is used in the profile likelihood, causing then that such profile likelihood becomes a function only of θ_1, as desired. The difficulty with this method lies in the fact that, when the number of nuisance parameters is high, the estimators obtained can be inefficient or even inconsistent. For this and severeal other reasons mentioned later on, Cox & Reid (1986) proposed using the conditional likelihood function, given the MLE of the orthogonalized parameters, θ_2. Let us then consider the following notation: let y be a $n \times 1$ vector of observations from a random variable Y, which has a pdf $f_Y(y;\theta)$, and where θ is a $1 \times p$ vector of unknown parameters. Then, let $L(\theta; y)$ and $l(\theta; y)$ denote the Maximum likelihood function and the log-likelihood function of θ

respectively. From this notation, we can now proceed to the following definition.

If $\theta = (\theta_1, \theta_2)$ where θ_1 and θ_2 partition θ, and $i_{\theta_i \theta_j}$ corresponds to the ij^{th} entry of the information matrix, then we define θ_1 to be orthogonal to θ_2 if

$$i_{\theta_s \theta_t} = \frac{1}{n} E\left[\frac{dl(\theta; y)}{d\theta_s} \frac{dl(\theta; y)}{d\theta_t}; \theta\right] = \frac{1}{n} E\left[-\frac{d^2 l(\theta; y)}{d\theta_s d\theta_t}; \theta\right] = 0$$

for all $\theta_s \in \theta_1$ and for all $\theta_t \in \theta_2$.

This definition can be extended to the case where more than two sets of parameters are orthogonal.

Some interesting properties inherent to orthogonality are that, if we suppose that $\theta = (\psi, \lambda)$ and ψ and λ are orthogonal, then:

1. if $\hat{\psi}$ and $\hat{\lambda}$ are the Maximum likelihood estimators of ψ and λ respectively, then they are asymptotically independent.

2. the asymptotic standard error of ψ is equal independently of either λ's value being known or unknown.

3. there might be some simplifications in the numerical calculation of $(\hat{\psi}, \hat{\lambda})$.

4. if we denote by $\hat{\psi}_\lambda = \hat{\psi}(\lambda)$ the ML estimate of ψ when λ is given, we also have that if λ and ψ are orthogonal, then $\hat{\psi}_\lambda$ varies only slowly with λ.

5. if $\hat{\psi}_\lambda = \hat{\psi}$ for all λ, then λ and ψ are orthogonal.

A.1 Construction of orthogonal parameters

Consider a parameter of interest ψ and a vector of nuisance parameters $\lambda = (\lambda_1, \lambda_2, ..., \lambda_q)$, which we want to orthogonalize from one another. Let ϕ be a function, such that

$\phi_i = \phi_1(\psi, \lambda)$ for $i = 1, 2, ..., q$, and suppose the likelihood function is in terms of the parameters $(\psi, \phi_1, ..., \phi_q)$, in other words,

$$l(\psi, \lambda) = l^*(\psi, \phi_1, \psi_2, ..., \psi_q),$$

then the first and second derivatives of the likelihood function have the form

$$\frac{dl(\psi, \lambda)}{d\psi} = \frac{dl^*(\psi, \phi)}{d\psi} + \sum_{r=1}^{q} \frac{dl^*(\psi, \phi)}{d\phi_r} \frac{d\phi_r}{d\psi},$$

hence we obtain

$$\frac{d^2 l(\psi, \lambda)}{d\psi d\lambda_t} = \sum_{s=1}^{q} \frac{d^2 l^*(\psi, \phi)}{d\psi d\phi_s} \frac{d\phi_s}{d\lambda_t} + \sum_{r=1}^{q}\sum_{s=1}^{q} \frac{d^2 l^*(\psi, \phi)}{d\phi_r d\phi_s} \frac{d\phi_s}{d\lambda_t} \frac{d\phi_r}{d\psi} + \sum_{r=1}^{q} \frac{dl^*(\psi, \phi)}{d\phi_r} \frac{d^2\phi_r}{d\psi d\lambda_t}. \tag{A.1}$$

By taking the expectation of equation A.1, we have

$$\begin{aligned} E\left[\frac{d^2 l(\psi, \lambda)}{d\psi d\lambda_t}\right] &= \sum_{s=1}^{q} E\left[\frac{d^2 l^*(\psi, \phi)}{d\psi d\phi}\right] \frac{d\phi_s}{d\lambda_t} \\ &\quad + \sum_{r=1}^{q}\sum_{s=1}^{q} E\left[\frac{d^2 l^*(\psi, \phi)}{d\phi_r d\phi_s}\right] \frac{d\phi_s}{d\lambda_t} \frac{d\phi_r}{d\psi} \\ &\quad + \sum_{r=1}^{q} E\left[\frac{dl^*(\psi, \phi)}{d\phi_r} \frac{d^2\phi_r}{d\psi d\lambda_t}\right] \\ &= \sum_{s=1}^{q} \frac{d\phi_s}{d\lambda_t} i^*_{\psi\phi_s} + \sum_{s=1}^{q}\sum_{r=1}^{q} \frac{d\phi_r}{d\psi} \frac{d\phi_s}{d\lambda_t} i^*_{\phi_r\phi_s}, \end{aligned}$$

where $i^*_{\psi\phi_s} = E\left[\frac{d^2 l^*(\psi,\phi)}{d\psi d\phi_s}\right]$, $i^*_{\phi_r\phi_s} = e\left[\frac{d^2 l^*(\psi,\phi)}{d\phi_r d\phi_s}\right]$ and $\sum_{r=1}^{q} E\left[\frac{dl^*(\psi,\phi)}{d\phi_r} \frac{d^2\phi_r}{d\psi d\lambda_t}\right]$ is equal to zero.

By definition, ψ and λ are orthogonal if

$$E\left[\frac{d^2 l(\psi, \lambda)}{d\psi d\lambda}\right] = 0;$$

which leads to the equation

$$0 = E\left[\frac{d^2 l(\psi,\lambda)}{d\psi d\lambda}\right] = \sum_{s=1}^{q} \frac{d\phi_s}{d\lambda_t}\left(i^*_{\psi\phi_s} + \sum_{r=1}^{q} i^*_{\phi_r\phi_s}\frac{d\phi_r}{d\psi}\right),$$

but since the Jacobian of the transformation from (ψ,ϕ) to (ψ,λ) cannot be zero, then $\frac{d\phi_s}{d\lambda_t}$ can't be zero. This finally leads us to the equation

$$i^*_{\psi\phi_s} + \sum_{r=1}^{q} i^*_{\phi_r\phi_s}\frac{d\phi_r}{d\psi} = 0\ ,$$

and thus the orthogonalizing equation, which determines the dependence between ϕ and ψ, is given by

$$\sum_{r=1}^{q} i^*_{\phi_r\phi_s}\frac{d\phi_r}{d\psi} = -i^*_{\psi\phi_s}.$$

Observe that, if ψ had a higher dimension, there is no guarantee, for instance in the case that $\psi = (\psi_1,\psi_2)$, that the condition $\frac{d^2\phi_s}{d\psi_1 d\psi_2} = \frac{d^2\phi_s}{d\psi_2 d\psi_1}$ holds. For this reason, global orthogonality of parameters cannot always be obtained.

Appendix B

Deviations from correctly specified models

Let us remember that the model proposed in Hutton & Solomon[25] is a mixture of an accelerated life (AL) model, and a Cox's proportional hazards (PH) model:

$$m(t) = \left\{ g_0(t)\, e^{\beta' z} G_0(t)^{\exp(\beta' z)-1} \right\}^{\psi} \left\{ f_0\left(te^{\gamma' z} \right) e^{\gamma' z} \right\}^{1-\psi},$$

where the survivor functions $F_0\left(te^{\gamma' z}\right)$ and $G_0(t)^{\exp(\beta' z)}$ correspond to the accelerated life model and the proportional hazards model respectively. F_0 and G_0 are the base-line survivor functions with corresponding densities f_0 and g_0 . Here z is the vector of fixed covariates, and let β and γ be the associated vectors of unknown regression parameters for the proportional hazards and accelerated life families of models. The interest lies in finding expressions for the parameters of interest, β and γ, while considering ψ as a nuisance parameter.

Let us remember from the Cox & Reid (1987) paper, that the orthogonalizing

equation between the parameters (ψ, ϕ) is of the form

$$\sum_{r=1}^{q} i^*_{\phi_r \phi_s} \frac{d\phi_r}{d\psi} = -i^*_{\psi\phi_s},$$

where $i^*_{\psi\phi_s} = E\left[\frac{d^2 l^*(\psi,\phi)}{d\psi d\phi_s}\right]$, $i^*_{\phi_r\phi_s} = e\left[\frac{d^2 l^*(\psi,\phi)}{d\phi_r d\phi_s}\right]$ and l^* is the log-likelihood function corresponding to ψ and ϕ. In the case of this problem the exact expressions of this orthogonalizing equations are not obtainable, hence the mixture function $m(t)$ is normalized and equated to the likelihood function

$$L(\psi, \beta, \gamma; \alpha, t, z) = \frac{m(t)}{\int_0^\infty m(t)dt};$$

where α, which is assumed as known, are the parameters of the base-line survival distributions.

B.1 Accelerated Life model true

Supose that the true model is the Accelerated Life model, which implies that $\psi = 0$, and that we have the independent failure times $t = (t_1, t_2, ..., t_n)$ and z_i a vector of $1 \times q$ fixed covariates for the i^{th} individual. Then, in order to calculate $\int_0^\infty m(t)dt$, we aproximate

$$m(t)e^{-\psi\beta' z}e^{-(1-\psi)\gamma' z}$$

with a first-order Taylor series about $(\beta, \gamma, \psi) = 0$, and find that

$$\int_0^\infty m(t)dt \approx (1 - \gamma' z + c_1\psi)e^{\psi\beta' z}e^{-(1-\psi)\gamma' z},$$

where $c_1 = E_{f_0}\left[\log\left\{g_0(t)/f_0(t)\right\}\right]$. Observe that this requires the assumption that β and γ are small.

From this, we obtain that the likelihood function is

$$L(\psi,\beta,\gamma;\alpha,t,z) \approx \prod_{i=1}^{n} \frac{m(t_i)}{\int_0^\infty m(s)ds} = \prod_{i=1}^{n} \frac{\left[g_0(t_i)G_0(t_i)^{\exp(\beta' z_i)-1}\right]^{\psi} [f_0(t_i e^{\gamma z_i})]^{(1-\psi)}}{1-\gamma z_i + c_1\psi};$$

which, with some calculations leads to the log-likelihood expression

$$\begin{aligned} l = l(\psi,\beta,\gamma;\alpha,t,z) \quad \approx \quad & \psi \sum_{i=1}^{n} \left[\log\{g_0(t_i)\} + (e^{\beta' z_i} - 1)\log\{G_0(t_i)\}\right] \\ & +(1-\psi)\sum_{i=1}^{n} \log[f_o(t_i e^{\gamma' z_i})] - \sum_{i=1}^{n} \log(1-\gamma' z_i + c_1\psi). \end{aligned}$$

From this equation, the first and second derivatives of $l(\psi,\beta,\gamma;\alpha,t,z)$ are found with respect to β, γ and ψ, but the expectations of the second derivatives are analitically intractable, and therefore, they are once again approached by first-order Taylor aproximations. By doing this, two sets of orthogonalizing equations are found, but one set is non-informative on the transformation between γ and β, since the corresponding normalizing constant does not include β. In order to simplify the expressions, the explanatory variables are orthonormalized, so that

$$\sum_{i=1}^{n} z_{is} = 0, \quad \sum_{i=1}^{n} z_{is} z_{ir} = \delta_{rs}, \quad r,s = 1,...,q,$$

where q is the dimension of the explanatory variables, and δ_{rs} is the Kronecker delta. Finally, from this simplifications the orthogonalization equation is expressed as

$$-c_2\psi\frac{\delta\beta_s}{\delta\psi} = (c_2+c_4)\gamma_s + c_2\beta_s;$$

where

$$c_2 = E_{f_0}[\log G_0(T)] \quad \text{and} \quad c_4 = E_{f_0}\left[T\dot{y}(T)\log G_0(T)\right].$$

A solution of such partial differential equation is

$$\beta_s = -\gamma_s \left(1 + \frac{c_4}{c_2}\right), \tag{B.1}$$

and from this it can be observed that when the proportional hazards assumption is used, when the true model is a life accelerated model, the coefficients are proportional to the first degree. Ultimatly, the expression can be further simplified to

$$\frac{\beta_s}{\beta_1} = \frac{\gamma_s}{\gamma_1} \text{ for } s = 2, 3, ..., 1,$$

which means that the ratios of regression coefficients are consistent. As a special case, when the true model is an accelerated life model and the analysis is performed under a proportional hazards model, but the base-line distribution is correct, (or so to speak, $f_0 = g_0$), then the regression coefficients have an even more simple expression,

$$\beta_s = -\gamma_s(1 - c_4).$$

B.2 Cox Proportional Hazards model true

For the case in which the true model is proportional hazards, but for the analysis it is assumed accelerated life model, the procedure is exactly the same as seen for the converse case. Once again, β and γ are assumed to be small, and after following the same steps around the value $\psi = 1$, we reach the orthogonalization equation

$$d_4 \left(\psi - 1\right) \frac{\delta \gamma_s}{\delta \psi} = d_4 \gamma_s + d_6 \beta_s,$$

where

$$d_4 = E_{g_0}\left[T\dot{y}(T) \log\left\{g_0(T)/f_0(T)\right\}\right] \quad \text{and} \quad d_6 = E_{g_0}\left[T\dot{y}(T)\left\{1 + \log G_0(T)\right\}\right].$$

A solution of such partial differential equation is given by

$$\gamma_s = -\frac{\beta_s d_6}{d_4}; \tag{B.2}$$

observe that this implies that the regression coefficients are linearly related under model misspecification. Furthermore, this relationship can be re-expressed as

$$\frac{\gamma_s}{\gamma_1} = \frac{\beta_s}{\beta_1} \text{ for } s = 2, 3, ..., q,$$

meaning that the ratios of the estimated regression parameters are consistent.

B.3 Inclusion of random censoring

Under the assumption of now having censoring in the data, the likelihood function is modified in order to include the censoring. The new orthogonalization equations are approximated by first-order Taylor series, with ψ, β and γ near 0. After calculating the first and second derivatives of the log-likelihood function, it is found that the orthogonalizing equations coincide with the orthogonalizing equations corresponding to the uncensored case. An analogous result is reached when it is considered $\psi = 1$.

It is concluded that the proportional properties found before still hold when right-censoring is included in the model.

B.4 Importance of Model selection

Observe that if the orthogonalizing equations (B.1) and (B.2) hold simultaneously, then it follows that

$$1 + \frac{c_4}{c_2} = \frac{d_4}{d_6}.$$

Furthermore, if the base-line distributions are the same, $g_0 = f_0$, then it $c_2 = -1$, $d_4 = c_5 - 1$ and $d_6 = c_4 - 1$, where $c_5 = E_{f_0}\left\{T^2 y''(T)\right\}$. Under such a case, the resulting expression

$$P = \frac{(1 - c_4)^2}{1 - c_5}$$

is proposed as a measure of how far are the hazards from proportionality. A value of P far from 1 would indicate a clearer departure from proportionality for the chosen model. This formula is proposed and named the proportionality factor.

Appendix C

Hutton & Monaghan[24] results

In this paper, the main objective is to study the properties of misspecification of two sorts, given that Cox's proportional hazards and accelerated life models are to be compared. As it was investigated in previous works (Hutton & Solomon), given that covariate coefficients are assumed small and that baseline parameters are considered known, and if a Cox's proportional hazards is misspecified as an accelerated life model, then the regression coefficients are consistent between the true and the assumed models. The result is shown to hold for the converse situation, in which the assumed model is a proportional hazards, and the true model is an accelerated life one. When orthonormalizing covariates, it was found that the importance of the covariates is preserved, although in a following work, Kwong & Hutton (2002), they suggest that orthonormalization of the covariates is not essential. An important difference from previous works, is that this paper does not assume that the regression coefficients are "small".

Three categories of misspecification are considered for this paper: when the base-line distribution is false, but the covariate effects are well specified, when the covariate effects are false but the right base-line distribution is considered, and finally when both covariate effects and base-line distribution are misspecified. The three distributions considered here for the base-line distribution are: log-normal,

log-logistic and Weibull distributions.

Suppose there are two treatments to be compared, which are the standard ($x = 0$) and the new ($x = 1$) treatment. If the true model is given by $f(t;\alpha)$ while the fitted model is $g(t;\beta)$, then for a sample of size n, $t_{1,...,t_n}$, the likelihood function will be

$$L_g = \prod_{i=1}^{n} g(t_i;\beta).$$

As quoted from Cox (1961) then the maximum likelihood estimator $\hat{\beta}$ under the misspecified model, will be asymtotically distributed as $N(\beta_\alpha, \frac{1}{n}C(\beta_\alpha))$, where β_α is the solution of the set of equations

$$E_f\left[\frac{\delta}{\delta\beta_j}\log g(T;\beta)\right] = 0 \text{ for } j = 1, ..., p$$

and

$$C(\beta_\alpha) = A^{-1}BA^{-1}$$

where A is the Fisher information matrix with $(j,k)^{th}$ element $-E\left[\frac{\delta^2}{\delta\beta_j\beta_k}\log L\right]$; and B has $(j,k)^{th}$ component $B(\alpha)_{jk} = E_f\left[\frac{\delta \log L}{\delta\beta_j}\frac{\delta \log L}{\delta\beta_k}\right]$. In order to calculate such derivatives, the log-likelihood was split into the failure times belonging to the standard an the new treatment. The expectation was taken on both sides of the equation and the asymptotic average was estimated for each treatment.

$$E_f\left[\log g(T)\right] = a_0 E_{f_0}\left[\log g_0(T)\right] + a_1 E_{f_1}\left[\log g_1(T)\right],$$

where g_0 and f_0 are the assumed and true distributions under the standard treatment, g_1 and f_1 are the assumed and true distributions under the new treatment and a_0 and a_1 are the proportion of observations for the standard and new treatmens respectively. In order to perform the desired asymptotic average estimation, the resulting equations were solved analitically if possible, but otherwise the Newton-Raphson procedure was carried out.

After all different types of misspecifications are considered and the parameters are estimated, a large series of detailed results are stated for every base-line distribution considered. Among the most important results found by this study are:

- The shape and regression parameters are biased when proportional hazards models are fitted to accelerated life models.
- For proportional hazards models, the direction of the bias also depends on the covariate distribution and effect size.
- For orthonormalized covariates, the regression coefficients are proportional (this was originally obtained in Hutton and Solomon (1997)).
- Consistently with Hutton & Solomon (1997), it was found that the effect of misspecification decreases as the centre of density of survival times move away from zero, which means that there are not too many early failures.
- When the fitted model is misspecified, the bias in the lower and upper percentiles is usually more substancial than the bias in the median.
- Fully parametric models give narrower confidence limits for quartiles than non-parametric methods, but they are subjected to the bias of assuming an incorrect model.
- The accelerated life model is more robust to misspecification because of its log-linear form.